MASTER THESIS
Master's Programme in Learning in Complex Systems
September 2015

Complex Human Behavior: Stimulus Equivalence and Eye-Tracking

Pedram Sadeghi

Faculty of Health Sciences
Department of Behavioral Science

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Abstract

One of the approaches towards understanding complex human behavior from the behavior analytic position is the stimulus equivalence paradigm. Understanding of stimulus equivalence is of great importance both in regards of conceptual and application purposes. Moreover, eye-tracking—procedures of using high technological equipment to obtain précis and accurate measuring of eye movements—is increasingly applied across of various scientific disciplines. The union of stimulus equivalence and eye-tracking could provide access and advancement in understanding of the nature of stimulus equivalence.

Article 1 is a conceptual piece presenting stimulus equivalence and the field of eye-tracking. The role of stimulus equivalences within behavior analysis is discussed and the different variables in which affect the establishment of stimulus equivalence classes is presented. Eye-tracking is introduced with a focus on, its historical highlights; application areas; the technology; different eye movement measures; and behavior analytic research that have employed eye-tracking technology.

Article 2 is an empirical study of stimulus equivalence in conjunction with the use of eye-tracking technology. The utilization of eye-tracking allows for the concept of observing response to be used to expand the investigation of how different training structures (i.e., linear series; many-to-one; the one-to-many) influence the establishment of stimulus equivalence class formation, which is the purpose of the study. Results with the additional measures are presented in the form of reaction time; fixation time; fixation rate; and transition rate, and discussed with respect to previous findings within stimulus equivalence research.

Keywords: Complex human behavior, stimulus equivalence, eye-tracking, eye movements, training structure, ocular observing response, linear series, many-to-one, one-to-many, attention, stimulus control.

Acknowledgements

First and foremost I would like to thank my supervisor Professor Erik Arntzen for his invaluable guidance and enthusiasm in this enterprise, and the opportunity to take on this subject matter. Special thanks to all the participants, the members of the lab group, all the people from the department of behavioral sciences that I have been fortunate to be surrounded by. I am grateful to the kind librarians at Kjeller, for without their assistancein allowing me to gain access to the complex and historically rich field of eye-tracking, it would very much remain novel and uncharted to me.

I am also immensely thankful to all the people close to me for their continuing support and unconditional love.

Article 1:

On stimulus equivalence and eye-tracking

Article 2:

The effect of training structures on stimulus equivalence and ocular observing response

List of Tables

Article 2:

The effect of training structures on stimulus equivalence and ocular observing response

List of Figures

Article 1:

On stimulus equivalence and eye-tracking

Article 2:

The effect of training structures on stimulus equivalence and ocular observing response

Complex Human Behavior:

Stimulus Equivalence and Eye-Tracking

On Stimulus Equivalence and Eye-Tracking

Pedram Sadeghi

Oslo and Akershus University College

Abstract

One of the approaches to understanding complex human behavior from a behavior analytic standpoint is the investigation of stimulus equivalence. This paper takes on what stimulus equivalence is, and variables that affect the establishment of stimulus equivalence. Moreover, since it has been suggested that the development and advancement of behavior analysis are intertangled with expanding the ways of measuring behavior, this paper also gives an introduction and description of the field of eye-tracking. Eye-tracking is the process of measuring where and how one is seeing within a visual scene. Although, eye-tracking is utilized in various fields of scientific disciplines, the technology is nevertheless rarely utilized within the field of behavior analysis. On the occasions when it has been employed it has been in conjunction with stimulus control, attending, and complex behavior. Finally, in this paper behavior analytic research in conjunction with eye-tracking is presented, where its utilization has been advantageous to the understanding of stimulus control relations.

Keywords: Complex human behavior, stimulus equivalence, eye-tracking, eye movements, attention, stimulus control.

On stimulus equivalence and eye-tracking

Contrary to the popular belief among those that are unfamiliar with or rather have an lack of understanding of the behavior analytic approach and its underlying philosophy (e.g., Brysbaert & Rastle, 2009; Chomsky, 1959), it is perfectly capable to undertake and does investigate complex human behavior, such as emergent or derived behavior (Chiesa, 1994; Donahoe & Palmer, 2004), which has shown to be of critical importance in the advancement of clinical applications (Dougher, Twohig, & Madden, 2014; Guinther & Dougher, 2015). Within behavior analysis this can be seen in the context of a question that has caused fascination and debate, which is the question of how an organism is prone to treat different events as if they were the same, especially when there has neither been a direct relation between the events, nor any physical similarities (Arntzen, 2010; Green & Saunders, 1998; Sidman, 1994). This phenomenon where two divergent events become interchangeable, equivalent, as a result of their relation to a third event that the organism has experienced in its environment, is described as stimulus equivalence. Within behavior analysis stimulus equivalence is more and more frequently called upon to describe various complex behavior, e.g., language, cognition, meaning, semantics, comprehension, symbols, symbolic behavior, and rules (Sidman, 1992). An example of this is how the stimulus equivalence paradigm has been used to teach children with intellectual disabilities academic skills such as geography and telling time, which had favorable results (Arntzen, Lian, & Halstadtrø, 2011).

There has been an ongoing focus of uncovering variables that affect stimulus equivalence formation (Arntzen, 2012). Fields and Verhave (1987) refer to hierarchy encompassing four parameters to completely describe the logical organization of stimulus equivalence class. The four parameters are class size; the number of nodes; distribution of singles among nodes; and directionality of training. Furthermore, there are three training structures, i.e., linear series (LS); many-to-one (MTO; also called comparison-as-node); and one-to-many (OTM; also called sample-as-node), and three training- and test protocols, i.e., simple-to-complex; complex-to-simple; and simultaneous, which affect outcome of stimulus equivalence. Other variables can be instructions; stimulus familiarity, i.e., stimuli that are correlated with some kind of learning history (Arntzen, 2012).

Dymond and Rehfeldt (2001) have proposed how additional measures can supplement research on derived stimulus relations, e.g., reaction time. Although additional measures are not directly a part of the stimulus equivalence definition, the utilization of the measurement has contributed to shed light on aspects of stimulus equivalence, which have had immense importance and impact on the line of research, e.g., Spencer and Chase (1996) and Pilgrim and Galizio (2000). Likewise, Palmer (2010) mentions eye-tracking as one of the methods one can use to measure behavior, which is seldom utilized. Eye-tracking is a procedure for measuring eye movements that has a long history dating back to 1800s and even further if considering contemporary research on oculomotor anatomy and vision (Wade, 2010; Wade & Tatler, 2005). Since, the technological advances in the 1960s and 1970s eye-trackers have become less invasive and more precise. Eye-tracking is being utilized in more and more various fields of research (Richardson & Spivey, 2004a). Other disciplines practice eye-tracking to study numerous of exactly same subject matters in which we strive to explain within behavior analysis.

The advancement of the field therefore hinges upon understanding of well-established concepts as well as an expansion of procedures that enable the uncovering of environment-behavior relations (Palmer, 2010). The purpose of this paper is fourfold, (1) describe one of the behavior analytic approaches to complex human behavior, i.e., stimulus equivalence; (2) account for some of the variables that affect stimulus equivalence, (3) give an introduction to what eye-tracking is, its history, and other core concepts as applications, equipment, and behavioral measurements; and (4) how eye-tracking has been utilized within behavior analysis. The scope of this paper does not permit an in-depth and detailed account focusing solely on the topic of eye-tracking history. Therefore turning points in history will be recounted, and additional historical facts incorporated into the subsequent subheadings so as to give a better understanding of their historical impact and importance.

Stimulus Equivalence

Within behavior analysis some of the first studies related to what was later to become stimulus equivalence research were described as early as in the mid-1930s (F. S. Keller & Schoenfeld, 1950/2011; Sidman, 1994), under terms such as mediated association (Peters, 1935); mediated generalization (Cofer & Foley Jr, 1942; M. Keller, 1943); and semantic conditioning (Riess, 1940, 1946). This type of research was concluded when scientists were not able to report consistent positive findings, and it was not until the 1970s that Sidman and colleagues revived the interest. Sidman, who was initially interested in reading comprehension, conducted his first stimulus equivalence study in 1971, where he studied the behavior of a 17-year-old boy. The study showed that derived relations (stimulus equivalence) were established between visual words and pictures, after each of them independently had become equivalent to auditory words (Sidman, 1971).

The research field expanded when Sidman and colleagues (Sidman & Tailby, 1982) published their article in 1982 (Arntzen, 2010). Research findings related to stimulus equivalence have both great conceptual implications as well as great implications related to practical application within behavior analysis (Arntzen, 2010). The conceptual implications are considering that some believe that stimulus equivalence is part of what makes the analysis of cognition feasible (i.e., Hayes, Barnes-Holmes, & Roche, 2001), and concerning the expansion of the behavioral unit, while the implications related to practical application are especially related to for instance education, where Sidman (1994) very clearly stated that he wanted more focus on applied research. Moreover, related to the clinical aspects of behavior analysis, stimulus equivalence research also contributes to advance what is known as the third-wave cognitive behavioral therapies, thereby making stimulus equivalence research important (Guinther & Dougher, 2015).

Sidman referred to *set theory,* where the mathematical definition of stimulus equivalence requires each stimulus-stimulus relation to hold the properties of reflexivity, symmetry, and transitivity. In stimulus equivalence research it is common to use a notation system, where numbers refer to class affiliation and letters to members of a class. Reflexivity holds that a conditional relation also needs to be in relation with itself. Reflexivity can be tested with generalized identity matching, e.g., A→A. B→B, and C →C. With the symmetry relation holds that the relationship must be bidirectional, i.e., when one has learned A→B, the A and B switch places, and symmetry is tested with B→A. Transitivity means testing a relationship that has not been explicitly taught, e.g., when A→B and B→C is learned, A→C is tested (Sidman, 1992, 1994). The stimulus equivalence test, also called the global equivalence test, i.e., a combined symmetrical transitivity relation, which can be tested when A→B and B →C is learned, and C→A is tested (Sidman, 1992, 1994). In this manner, we can uncover if a derived relation has emerged. To be able to respond correctly to the equivalence test, both directly trained relations must hold the properties of equivalence. If only one of the individual properties is absent, one could expect a negative equivalence test (Pierce & Cheney, 2008; Sidman, 1992). Sidman points out that the equivalence relation only can be observed in the sense that we can make inferences from the results of the tests. If stimulus equivalence test show derived responding, this means that this is a case of equivalence class, where all stimuli are interchangeable (Sidman, 1992).

It is critical to have an understanding of conditional discrimination procedure to study stimulus equivalence, since a conditional discrimination procedure as with arbitrary- or identity matching-to-sample (MTS) is a fundamental procedure in this type of research

(Saunders, 1989; Sidman, 1994; Sidman & Tailby, 1982). In a conditional discrimination procedure one sample stimulus is presented with two or more comparison stimuli, where correct responding is reinforced. It is called arbitrary MTS since there are no physical similarities between the stimuli (Sidman, 1992). In addition, there is a variation of MTS called *delayed matching-to-sample* (DMTS), where the sample stimulus is removed before the comparison stimuli are presented (Blough, 1959). The delay can be from 0 s and longer.

Sidman and colleagues also pointed out that even though MTS and conditional discrimination are procedurally identical, results would differ as one produces conditional discrimination and the other true matching to sample. True matching to sample represents a semantic process (i.e., semantic correspondence), which conditioned discrimination does not. When there is true machining to sample, the defining class will make all samples and their corresponding comparison equivalent to each other (Sidman, 1994; Sidman et al., 1982).

It is recommended that it should be at least three comparisons stimuli, so as to restrict the possibility of responding in a manner that excludes stimuli, so-called ”reject control” (Carrigan & Sidman, 1992; Johnson & Sidman, 1993; Sidman, 1987). Also, with *two-choice* MTS there is a 50% probability of responding correct, which could result in misleading findings (Green & Saunders, 1998).

Through an MTS procedure, a participant learns to match a sample stimulus, based upon which comparison is present. Thereby the response comes under the control of a conditioned stimulus. This can be described with the four-term contingency Sc – Sd: R – Sr, and is often termed contextual discrimination, since the three-term contingency comes under control of a conditional stimulus (Sc) (Sidman, 1994). Furthermore, Sidman (1994) refers to how units of behavior can be termed as five-term contingencies; in a five-term contingency a four-term contingency comes under the control of a conditional control. Even though Sidman initially considered a four-term contingency a requirement for establishing stimulus equivalence, later he stated that a two-term contingency could be sufficient.

Variables Affecting Equivalence Class Formation

Fields and Verhave (1987) describe the four parameters of (1) class size, i.e., number of stimuli in a class; (2) number of *nodes,* i.e., the number of stimuli that each are linked through training to at least two other stimuli; (3) distribution of *singles* among nodes (i.e., nodal density), the number of singles linked through training to each node (i.e., Training cluster)—a single is a single stimulus linked through training to only one other stimulus; and (4) directionality of training, i.e., the manner of which stimuli in a class are presented as sample stimulus and comparison stimuli, to completely describe the structure of stimulus equivalence classes. Furthermore, they point out that the values for each parameter has to be set in this hierarchical order described above.

The class size determines the maximum number of derived relations as well as limits all other structural parameters of an equivalence class. The manner in which stimuli are linked is either unidirectionality, e.g., A→B and B→A or bidirectionality A←→B.
In their article they use the term *associative distance* (i.e., nodal distance), which is the number of nodes between two singles.

Three training structures and three training- and test protocols are utilized to uncover how derived relations are established and how variables affect responding in accordance with stimulus equivalence. The three fundamental training structures used to establish conditional discrimination are: (1) LS, where A→B is trained, then B→C, so C→D, thereafter D→E, for so to mix some of the relations before testing for derived relation have emerged, e.g., E→A; (2) MTO, where different sample stimuli are trained to one fixed comparison stimuli within a class so that B→A, C→A, D→A, and E→A relations are trained before being tested for derived responding, e.g., E→B, in the test half of the stimuli used as sample stimulus are removed and half of the comparison stimuli are presented as sample stimulus; and (3) OTM, where a fixed comparison stimulus is trained to different comparison stimuli, e.g., A→B, A→C, A→D, and A→E are trained before testing for derived responding, e.g., E→B. In relation to testing and the OTM training structure, the initial sample stimuli are removed and half of the comparison stimuli are presented as sample stimulus (Arntzen, 2010; Arntzen et al., 2011; Green & Saunders, 1998). There are no explicit rules when deciding which of the training structure one should utilize, but Green and Saunders (1998) refer to some general guidelines, where the most important factor is that the test protocol should be in line with the study objectives in a logical manner. Still there have been some inconsistencies in reports with respect to which training structures are most effective in producing responding that is in accordance with stimulus equivalence.

Three training- and test protocols utilized are (1) simple-to-complex where all relations are trained and tested before a global equivalence test (e.g., symmetry test); so transitivity test; then global equivalence test), (2) Complex-to-simple, which involves the opposite; and (3) simultaneous, where all relations are presented from start (Arntzen, 2010; Green & Saunders, 1998; Imam, 2006). Furthermore, Arntzen (2010) refers to the use of the simple-to-complex protocol to have been very effective in producing positive results on equivalence test, and Green and Saunders (1998) refers to other researchers who have raised the issue of how discrimination between potential stimulus classes can be incomplete when sample stimuli are nodes (OTM). Finally, it should be mentioned that other variables affecting stimulus equivalence could be, response requirements (e.g., observing response); whether it is SMTS or DMTS; the degree of detail given in the instruction; the type of stimuli used, whether it is abstract stimuli or meaningful stimuli; and of course participant characteristics as age and those with a medical diagnosis (Arntzen, 2012).

Eye-tracking

What is Eye-tracking

Eye-tracking is the process of measuring eye movement to identify where a person is looking, i.e., the absolute point of gaze in the visual scene, and how a person is looking (Bojko, 2013; Majaranta & Bulling, 2014). Generally speaking Hayhoe (2004) states that Eye-tracking allows us to measure for instance the fixation location, duration, onset, and accuracy, which often will allow stronger inferences then other measures, e.g., percentage correct or reaction time.

Application Areas

Eye-tracking is widely utilized in the fields of neuroscience; cognitive psychology; psycholinguistics; psychiatry; ergonomics; advertising; design; medical diagnostic; human-computer interaction; in-vehicle research; sports; infant research; and usability research (Bojko, 2013; Hayhoe, 2004; Holmqvist et al., 2011; Majaranta & Bulling, 2014; Rayner, 1998; Richardson & Spivey, 2004b).

Eye-tracking has been applied to assess how different drug labels affect how pharmacists attend to visual information (Bojko, Buffardi, Lew, & Israelski, 2006; Bojko, Gaddy, Lew, Quinn, & Israelski, 2005); to explore reading comprehension of first and second language readers (Kang, 2014); to study the relationship between food/beverage advertisement and unhealthy choices of children and adolescents (Velazquez & Pasch, 2014); and to examine how infants acquire language (Ferguson, Graf, & Waxman, 2014).

History of Eye-tracking

The inescapable fact that the understanding of the ocular system has been of immense importance are evident as even Charles Darwin (1859/2006) referring to the eyes as the most perfect organ—which was of an extreme level of perfection and complication—confessed that claiming that the eyes were a product of natural selection sounded highly absurd, although he still argued that it was the case. While the study of the anatomy of the ocular system can be dated back to the medical research of Claudius Galen (ca 130–200)—which was relied upon over 1000 years later, in which Hunain ibn Is-hâq (ca 807–877) later translated into Arabic—the study of eye movements can be traced as far back as to that of Aristotle (ca 384–322 BC). Aristotle, concerned with binocular eye movements, had the belief that the eyes operated in unison, and moreover, he differentiated between convergence and version movements. His beliefs were not based on experiments he had conducted but on acute observations.

Later the influential Ibn al-Haytham (Alhazen) (ca 965–1039), also interested in binocular eye movements, was one of the first to realize that eye movements could be examined by observations rather than to rely on introspection. Nevertheless, many centuries were to pass before this was actualized, and indeed the advances were driven by theory rather than observations (Wade, 2010).

William Charles Wells (1757–1817), which was interested to see if conjoint eye movements were learned or innate behavior (Wells, 1792), introduced afterimages (photogene) (see Figure 1) to assess if the eye movements were voluntary or involuntary. Although, others have suggested otherwise, there are strong indications that Wells was indeed the first to conduct systematic studies on eye movements (Wade, 2010; Wade & Tatler, 2005). Furthermore, due to technical limitations, the research of the late 1800s had a tendency to focus more on the orientation of the eyes, opposed to focusing on eye movements in themselves. However, despite technical limitations a series of laws were formulated during this period, e.g., Hering's law; Donders' law; and Listing's law (Wade, 2010).

In conjunction with the topics of psychology and visual perception in the early 1900s—the behaviorist revolution was successful by exclusively focusing on public (overt) events, but the stimulus-response paradigm was restricted in regard to accessibility for the study of visual perception—the behaviorists focus was solely on eye movements and consequently when it was neglecting, the study of visual perception was left to the Gestaltist (Wade & Tatler, 2005).

When considering the evolution of the eye-tracking, there have been multiple moments where technological advances and breakthroughs have had a revolutionizing impact. Eye-tracking data were initially acquired through introspection or by experimenter's subjective observations of participants' eyes when using devices like afterimages, mirrors, telescopes, and peepholes. To counter the vulnerabilities of these subjective inquiries there was a necessity for objective measures (Richardson & Spivey, 2004b; Wade & Tatler, 2005)

Eye-trackers

An eye-tracker is the device used for measuring eye movements. One of the major turning points came when Dodge and Cline (1901), invented the first noninvasive eye-tracker the "Dodge Photochronograph", which inspired further development (Wade & Tatler, 2005). Duchowski (2007) referrers to four broad categories of methods for measuring eye movements, electrooculography; scleral contact lens/search coil; photo-oculography or video-oculography; and video-based combined pupil/corneal reflection. There are primarily two types of eye-trackers, the first type measures the position of the eyes relative to the head, and the second type that measures the orientation of the eyes in space, which is called "point of regard" (Young & Sheena, 1975). To obtain point of regard measurement, the head must be stationary or several features of the eyes must be measured to distinguish between head movements and eye movements, e.g., the pupil center and corneal reflection (Duchowski, 2007).

Electrooculography relies on measurement of the skin's electric (corneoretinal) potential differences, in the form of a d.c. signals, by placing electrodes around the ocular cavity. These were widely utilized in the 1970s (Duchowski, 2007; Young & Sheena, 1975) (see Figure 2). Unless electrooculography is used in combination with a head tracker, is not suitable for measuring point for regard (Duchowski, 2007).

Scleral contact lens/search coil methods are based upon connecting a mechanical or optical reference object to a contact lens, which is directly placed on the eye. The most basic method is to measure a wire coil, which moves through a magnetic field. To avoid slippage, the lens has to cover the entire cornea and sclera (Young & Sheena, 1975). Even though this is one of the most précis methods of measuring eye movements, it can be very invasive for the participant and demanding in ways of implementation. Both of the above-mentioned methods are generally not suitable for point of regard measurements.

Photo-oculography or video-oculography methods refer to a variety of different eye-trackers, which generally do not measure point of regard (see Figure 3). They involve the measurement of identifiable features of the eye, e.g., pupil shape, the position of the limbus and corneal reflection with a nearby a light source (often infrared), during rotation/-translation. The measures provided by these eye-trackers may be obtained by automatic registration or via visual inspections of video recordings.

Eye-trackers that are video-based combined pupil/corneal reflection make use of cameras and image processing hardware to obtain the point of regard in real-time. Currently the most commonly used eye trackers are those which are based upon measuring corneal reflection (Purkinje reflections) relative to the pupil center (Duchowski, 2007).

Different Eye Movement Measures

Besides the development of the first noninvasive eye-tracker, one of the most critical breakthroughs within eye-tracking was the discovery of saccades. Although it was assumed that eye movements were quick, smooth, and continuous, those concerned with the subject of vision had observed the rapid eye movements that we today call saccades. Since they were restricted from the precise measurement that we today have access to, the technical restrictions limited the experimental studies to the orientation of the eye rather than eye movements (Wade, 2010; Wade & Tatler, 2005). It was Louis-Émile Javal (1839-1909) who introduced the term saccade, but his recognition of them and the necessity of objective measures were not based upon any objective measures, but rather his assertions based on his intuitions and subjective impression of eye movements. Javal was interested in vision during reading, and even though he made numerous attempts of measuring saccades with the use of afterimage or by measuring the deflections of light from a mirror attached to his eye, he was not successful (Wade & Tatler, 2005). In 1879 M. Lamare, a student of Javal was the first to observe and record saccades during reading, but it was not until 13 years later that he described his experiment (Wade, 2007). Lamare used multiple methods to measure the discontinuity of eye movements. One method was calculating numbers of words he could read divided by estimated numbers of pauses made, another method was to count numbers of distinct eye movements while looking along a line (Wade & Tatler, 2005). Both Lamare and Karl Ewald Konstantin Hering (1834–1918) measured saccades, with the use of two rubber tubes functioning like a miniature stethoscope, which were placed on the eyelids before listening to the sounds of the ocular muscles. Hering observed clapping sounds when the participants were reading, but the sound ceased when participants were instructed to fixate on stationary stimuli. He concluded that the sounds were of contracting oculomotor muscles that accompanied the eye movements. With the use of an afterimage, Hering studied the correlation of the clapping sounds and the movements, and thereby provided evidence for the conclusion (Wade, 2007). Hering described the discontinuity of eye movements and identified the class of rotations, which we now refer to as saccadic, but due to Huey´s less careful phrasing in his book, *The psychology and pedagogy of reading* published in 1908, Javal was credited with the discovery and measurements of saccades. Javal, had only made reference to M. Lamare’s unpublished work in a footnote, but in fact it was Lamare and Hering, which were the first to measure saccades, and this with very similar methods (Wade, 2007, 2009, 2010; Wade & Tatler, 2005).

For Dodge (1907) the relationship between eye movements and fixation was complexly related. He meant that the irregular eye movements that were regarded as accidental variations of fixations, were a part of the premise that justified the assertion; that absolute fixation did not exist within normal vision. He later described how most common eye movements were comprised of alternations between series of rhythmic jerks and periods of fixation.

Saccades are a part of human species phylogenic endowment, it seems that evolution has selected the behavior of inspecting small portions of the visual world in rapid sequence, rather than the cumbersome alternative of devoting recourses to processing the enormous amount of visual data in detail (Richardson & Spivey, 2004b). Still it is only during fixations visual information attained, while an eye movement gives the eyes access to fixate on multiple areas in the visual scene.

As technological advances came about more inconspicuous eye movements were discovered. There are various types of eye movements, some of the more common ones are. e.g., *fixations*; *saccades*; *glissades*; *smooth pursuit*; *microsaccades*; *tremors*; *drifts*; and *scanpaths.*

Fixations. When measuring fixation we measure the eye when it "is rather fixed", however the eyes are never completely still. Dodge correctly noted that the "term 'fixation' is frequently used to cover the entire process of visual adjustment, including the antecedent eye movements. In general, however, fixation appears to mean that the point of regard remains relatively unchanged within the visual field." (Dodge, 1907, p. 1); Holmqvist et al. (2011) differentiate between the oculomotor definition, which Dodge (1907) talks about, and the processing definition that include visual intake as an additional criterion on fixations. Since, fixation durations in truth are calculated by various fixation detection algorithms (e.g., I-DT), which do not consider visual intake and have different definitions of fixations, fixation durations are exclusively defined by the event detection algorithms and their properties. A fixation typically has a duration of 200–300 ms, yet they can have a duration of down to 100 ms. Therefore the minimum duration threshold for dispersion-based identification algorithms is 100-200 ms (Salvucci & Goldberg, 2000). The function of fixations is to bring stability to the eye so that the stimulus is brought to the area of the retina called the fovea (Wade & Tatler, 2005). Fixation duration, also called “fixation time”, “dwell time”, or “dwell time of the fixation,” is probably the most used measure within eye-tracking research (Holmqvist et al., 2011).

Saccades. Generally speaking saccades can be defined as a period when the eye “moves fast,” they have a duration of 30–80 ms, amplitude of 4–20°, with the velocity of 30–500°/s (Holmqvist et al., 2011). Richardson and Spivey (2004a) define saccades as large ballistic scanning movements that occur 3–4 times per second. They are voluntary eye movements, with the function of bringing the fovea to the target of fixation. The saccades have characteristics of start with an initial acceleration, which is very high, and a deceleration at the end (Young & Sheena, 1975). Saccades are the fastest movements an external part of the body can produce. During a saccadic eye movement, vision sin suppressed and the sensitivity levels equals blindness (Bojko, 2013; Holmqvist et al., 2011).

Glissades. According to Holmqvist et al. (2011) glissade is a post-saccadic eye movement, where the eye “wobbles” a little before stopping up. A great amount of all saccadic eye movements, i.e., between 20–40%, end with a glissade, however nearly no saccade begins with a glissade. Glissades have a duration of 10–40 ms, an amplitude of 0.5–2°, with the velocity of 20–140°/s. Up till now various event detection algorithms have treated glissaded unsystematically and different, by sometimes considering them fixations and other times saccades, this sometimes within the same algorithm. Considering that more algorithms are becoming able to detect and treat glissades, there certainly will be more research that will utilize glissades in future research. Furthermore, Holmqvist et al. (2011) consider glissades as saccadic in nature, because of the lack of apparent visual intake, and that they follow the same main sequence as saccades.

Smooth pursuit. Smooth pursuits are slower eye movements that occur when the eyes are visually tracking a moving stimulus (Duchowski, 2007; Holmqvist et al., 2011). They have a velocity of 10–30°/s, and seem to be restricted in acceleration as well as in velocity. Smooth pursuits require a moving stimulus and the ability of the eyes to match the velocity of the moving stimulus (Holmqvist et al., 2011; Young & Sheena, 1975). Nevertheless, some studies have shown to contradict the requirement of moving stimulus. It is this that distinguishes smooth pursuits from saccades, where the latter can be made without the requirement of stimulus being present (Holmqvist et al., 2011). According to Holmqvist et al. (2011) in addition to being completely different movements, saccades, and smooth pursuit are controlled by different parts of the brain. Their function is to stabilize the image of the moving stimulus or background on the retina—this independent of the saccadic eye movements—and are primarily not under voluntary control (Young & Sheena, 1975).

Intra-fixational eye movements. There are three involuntary types of micro-movements during fixation, i.e., microsaccades; tremors; and drifts (Yarbus, 1967). These eye movements help to keep the stimulus in the center of the fovea, and prevent sensory adaptation in our visual path, which if not would leave the eyes blind during visual fixation (Martinez-Conde, Macknik, & Hubel, 2004).

Microsaccades. Microsaccades usually occur when the fixation duration has exceeded a certain length (0.3–0.5 s), or when drifts move the point of fixation too far away from the center of the fovea (Yarbus, 1967). It is the microsaccades function to rapidly bring the eye back to its original position, and have a duration of 10–30 ms, amplitude of 10–40', with the velocity of 15–50°/s (Holmqvist et al., 2011). The role of microsaccades are still uncertain in relation their influence on the maintenance of visibility (Martinez-Conde et al., 2004).

Tremors (physiological nystagmus). The fact that one must take into account, not only the movement of the participant's head, but also vibrations from the apparatus, in addition the building itself, makes the tremors the most difficult eye movement to measure. Tremors have very low amplitude and very high frequency (Yarbus, 1967). The exact role of tremors are unclear, it is possible that they are the cause of imprecise muscle control. Tremors have an amplitude of < 1', with the velocity of 20'/s (peak), and a frequency around 90 Hz (Holmqvist et al., 2011).

Drifts. Drifts are irregular and relatively slow eye movement moving the eyes away from the center of fixation, which occur between periods of microsaccades, and are always accompanied by a tremor (Holmqvist et al., 2011; Martinez-Conde et al., 2004; Yarbus, 1967). According to Holmqvist et al. (2011) drifts have a duration of 200–1000 ms, an amplitude of 1–60', with the velocity of 6–25'/s.

Scanpaths Noton and Stark (1971a, 1971b) initially introduced the term Scanpath, referring to a specific participants characteristic viewing pattern, made up of a fixed path while viewing a pattern, which appeared in the learning phase and reappeared at the beginning of the recognition phase. Still consistencies were not found related to a common characteristic in these patterns across of participants (Noton & Stark, 1971a, 1971b). According to Holmqvist et al. (2011) The term scanpath is currently used to describe how concrete eye movements come about physically through space, but this not necessarily for one participant and "define a scanpath as the route of oculomotor events through space within a certain timespan." (Holmqvist et al., 2011, p. 590).

Behavior Analysis and Eye-tracking

Among some of the research areas which have employed eye-tracking within behavior analysis, have concerned operant control of observing response; eye movements in relation to stimulus control with variations of DMTS; transfer of stimulus control; eye movements in relation to stimulus overselectivity; and how complexity MTS task affected observing. Schroeder and Holland conducted some of the earliest research concerning eye movements and observing response, with the use of noninvasive eye-trackers (Holland, 1957, 1958; Schroeder, 1970; Schroeder & Holland, 1968a, 1968b, 1969).

The observing response is a response in which emitting produces discriminative stimuli. The observing response is not reinforced in itself but is rather reinforced by secondary reinforcement. It is the response to the produced discriminative stimuli which leads to reinforcement that maintains the observing response, which produced the discriminative stimulus (Wyckoff, 1952, 1959). Since, the function of eye movements is to give access to information (Sd), and as Palmer (2010) asserts eye movements rarely are reinforced in itself, the measures of eye movements mentioned earlier can be classified as *ocular observing responses*. Furthermore, according to Dinsmoor (1985) there is a close relation between that of observing response and stimulus control, in that it is a prerequisite of efficient visual discrimination.

When Holland (1957) studied human observing response, he found that they were under operant control. This to the extent that responding was analogous to that of which characteristic animals response patterns with that of scallops when reinforcement schedule of fixed interval were in effect. Subsequently Holland (1958) employed different types of reinforcement schedules to assess if signal detection could function as a reinforcer. The results were positive; he found that the rate of observing was relative to the rate of signal detection (reinforcement). With the use of a head mounted eye-tracker (corneal reflection), Schroeder and Holland (1968a) replicated Holland (1958) findings that eye movements came under operant control. Furthermore, they proposed that measurement of eye movements could enable a more direct assessment of attention and that “these precurrent responses may account for some or all of the apparent stimulus selection in simple and complex discrimination learning tasks.” p.161. Schroeder and Holland (1968b) reported finding a relation between that of higher rates of eye movements and signal detection. Subsequently, a year later Schroeder and Holland (1969) reconfirmed their findings, and with the use of a concurrent reinforcement schedule and a change-over delay showed that participants matched relative rates of eye movements relative to the rate of reinforcement. Schroeder (1970) set out to investigate the extent to which fixations corresponded to correct responding in a simple discrimination task and that of the effect of repeated exposure (i.e., practice effect) to the task had on fixations. The results were that all the participants fixated most on stimuli correlated with reinforcement, and that there were least fixations on the stimuli in which was not responded to; and that repeated exposure to the task lead to a reduction in fixation on the stimuli, where there was faster decrement in fixation of stimuli responded upon.

Schroeder and Holland (1969) concluded by suggesting that, “these results, together with the present results, suggest that operant control of eye movements plays an important part in establishing stimulus control and can be a powerful tool for assessing functional properties of stimuli.“ p. 903.

More recently Pessôa, Huziwara, Perez, Endemann, and Tomanari (2009) replicated the Schroeder (1970) study and expanded it by balancing the luminance of the stimuli and evaluating the effect it had upon the practice effect. Their findings were that correctly chosen stimuli were both responded upon more often and were significantly more fixated upon relative to other stimuli. However, whereas Schroeder (1970) results showed no fixations at the end of the experiment, Pessôa et al. (2009) data showed that although a decrease in fixation duration, responding did not occur without the minimum of one fixation upon the chosen stimuli. Their conclusion was that balancing the luminance of the stimuli could increase the probability that stimuli will be fixated upon and thereby reducing practice effects, i.e., increase the fixation duration for target stimuli.

Also Dube and colleagues (Dube et al., 1999; Dube et al., 2003) have employed eye-tracking. This when investigating stimulus overselectivity in multiple-samples (two samples) DMTS tasks (IDMTS), in individuals with intellectual disabilities. Eye-tracking data obtained from one participant showed that on trials where only one sample stimulus was observed, responding would be correct if the sample was the correct comparisons, while responding would be at chance level if it was the incorrect comparison. Thereby the data indicated that stimulus overselectivity could be a product of failure to observe all of the relevant stimuli, i.e., emitting the appropriate observing response.

Dube et al. (2010) expanded the previous study (Dube et al., 1999; Dube et al., 2003) with 13 more participants. Thus studying 14 participants whereof four were normally capable adults and 10 participants with intellectual disabilities attending residential schools. In addition to assess the relationship between observing response and two sample DMTS (IDMTS); they wanted to assess those participants showing intermediary accuracy and lack of observing response in the initial evaluation; and if experimental procedures increasing observing durations could produce higher accuracy and thereby abolish stimulus overselectivity. Five participants showing intermediary accuracy were given either differential reinforcement, extra-, or within-stimulus prompts. If the intervention was unsuccessful in producing increased observing behavior and higher accuracy, additional interventions either alone or in combinations, were implemented. The interventions consisted of differential reinforcement for observing; extrastimulus prompts; within-stimulus prompts; observing contingency; within-stimulus prompts plus observing contingency; and high-accuracy contingency. They conclude that at least some instance of stimulus overselectivity were caused by lack of appropriate observing behavior and that this it partially correctable with behavioral interventions.

Dube et al. (2006) examined how task complexity affected observing behavior with the use of multiple-sample DMTS (IDMTS) and eye-tracking. By comparing the presentation of two or four sample stimuli per trial, they found that participants with high accuracy exhibited higher observing duration relative to participants with low accuracy. Furthermore, they found that for participants with low accuracy, the transition to high accuracy brought about a slight increase in the observing frequency for one of the participants, while there was a considerable increase in observing duration for both of the participants. They concluded that the data suggested that different aspects of observing behavior topography could be independent.

Tomanari et al. (2007) expanded on the Dube et al. (2006) study and examined the relationship of manual observing response and ocular observing response during a Wyckoff observing response procedure. The results were that prior to reaching the discrimination ratio requirements for the manual observing response, there was an acceleration in the rate of ocular observing response for three of the four participants who completed the experiment. They also found that relative to the manual observing response, all participants emitted more frequent ocular observing response that were of shorter duration. The observing durations of the S+ and S- where alike for the manual and ocular observing response. Thereby the manual observing response predicted the ocular observing response. In addition, they found that when there was a response requirement to produce the stimuli, the S- was observed for a longer duration relative to S+ while the opposite was seen when the stimuli were continuously presented. In conclusion their findings were similar to those of Dube et al. (2006) that the two observing responses had very different response topographies. They attributed the differences to the possibility of being caused by the response cost. In which the ocular observing response is much more cost efficient compared to moving hand and arm to respond.

With the purpose of replicating previous research findings, such as Reynolds (1961) seminal research concerning "attention" in pigeons, Perez, Endemann, Pessôa, and Tomanari (2014) used eye-tracking and different stimulus-control assessment procedures to assess how human participants exposed to discrimination training with compound stimuli (color or shape) responded during tests. In sum, there were two phases (reversed conditions/contingencies), each consisting of one training and one testing phase, where the testing phase consisted of six different test where compound stimuli separated and/or recombined to form novel stimuli. In test 1, all the components were separated and then presented; in test 2–5 each of the S+ components was separated and pitted them against an S- component; and test 6 where components recombined into a novel stimuli. Results from both phases where that all the participants ocular observing response towards one component of the S+ was of higher frequency and fixation duration, relative to the other S+ component and two S- components. Also the second highest frequency and fixation duration were allocated to the S- of the same dimension as S+, which the participant showed most ocular observing response towards. All participants in test 1, except for one, responded to the component with most ocular observing responses. Conversely across tests 2–5 participants always responded to the S+ component, even though this was not the component they observed most during training. Subsequently in test 6 all participant, expect for one in phase 1, again responded to the component with most ocular observing responses during training. Therefore, they concluded that although participants tend to respond to the stimulus component that correlated with most ocular observing responses during training, there were idiosyncratic difference among the participants towards which component type they responded on the basis of, and that S+ component that was correlated with less observing responses also affected participants responding during tests (i.e., rejection control). Furthermore, they reconfirmed the possibility of utilizing eye-tracking to assess stimulus control relations as those mentioned above (Perez et al., 2014).

Kirshner and Sidman (1972) also studied ocular observing response, this in the form of scanning of a visual display, for three aphasic patients performing MTS tasks. They found (1) that different sample stimuli (auditory spelled stimuli) changed scanning behavior, (2) that there where higher number of observing response on correct comparison stimuli on the tasks which they showed high error rates, (3) that there was change in scanning on tasks they earlier hade high error rates, which later they during testing responded correctly on and (4) that when comparing error scores and scanning they found that ocular observing response was more precise in detecting language difficulties.

Ultimately, Madelain, Paeye, and Darcheville (2011) posed the question of if the eye movement measures of saccades and smooth pursuit are operant behavior and concluded that they indeed are. They found that dimensions like velocity, latency and reaction time are the under the control of reinforcement contingencies.

Conclusion

The purpose of this paper has been to direct attention to complex human behavior in the behavior analytic sense, with the aim of proposing additional measures of assessing complex human behavior, which has been utilized in various other fields. Although eye-tracking nor any other measuring instrument is neither *a* solution nor *the* solution in itself, the research from both within and outside of behavior analysis shows that there is much to be gained by implementing the technology. Behavior analytic research with eye-tracking has shown that eye movements are under operant control and given more a precise measurement when assessing stimulus control. Moreover, it clarifies that the three-leveled model of selection as described by Skinner (1981), which behavior analysis is founded upon are continuous even on microscopic levels, i.e., behavioral units of incredible small size as the eye movements discussed above. The aforementioned entails that not only is eye-tracking a valuable resource but that the behavior analytic approach also can be fruitful for others who conduct work within the field of eye-tracking that do not have a behavior analytic background. In addition, the field of eye-tracking has had an equivalent development as that of psychology in the sense that both independently reached a conclusion of critical importance, i.e., the role of objective measures, which is the pinnacle of natural science.

Their lies an excellent opportunity to reap the benefits of eye-tracking by utilizing it in combination with stimulus equivalence research—which never has been published before—where a further line of research could investigate variables in which have shown to have an effect on the establishment of stimulus equivalence class formation, but no complete conclusion has been made available. A perfect example of this is the case of how training structures affect the establishment of stimulus equivalence class formation, where slightly different findings have made it difficult to reach an end conclusion.

Next, the consequence of filling the temporal gaps where behavior usually is only privileged to one person—while possibly not being aware of their own behavior—with reliable measures would give a major insight into making covert behavior, overt behavior. Thereby possibly strengthening the position of behavior analysis as a hard science and improving behavioral applications. On a general note, if a stronger consilience was to be reached between the fields of science, it could be the case that in long term it would contribute to reaching the acme of knowledge when considering stimulus control and thereby contributing to the improvement of human life.

References

Arntzen, E. (2010). Stimulusekvivalens. Teoretiske betraktninger og noen praktiske implikasjoner. In S. Eikeseth & F. Svartdal (Eds.), *Anvendt atferdsanalyse 2 ed.* (pp. 100–138). Oslo: Gyldendal Akademika.

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*, 123–135. Retrieved from http://www.ejoba.org.

Arntzen, E., Lian, T., & Halstadtrø, L.-B. (2011). Anvendelse av matching-to-sample prosedyrer i etableringen av akedemiske ferdigheter. *Norsk Tidsskrift for Atferdsanalyse, 38*, 1–26 Retrieved from http://www.nta.atferd.no.

Blough, D. S. (1959). Delayed matching in the pigeon. *Journal of the Experimental Analysis of Behavior, 2*, 151–160. doi: 10.1901/jeab.1959.2-151

Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. Broklyn, NY: Rosenfeld Media.

Bojko, A., Buffardi, K., Lew, G., & Israelski, E. (2006). Eye tracking study on the impact of the manufacturer's logo and multilingual description on drug selection performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*, 1112–1116. doi: 10.1177/154193120605001024

Bojko, A., Gaddy, C., Lew, G., Quinn, A., & Israelski, E. (2005). Evaluation of drug label designs using eye tracking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49*, 1033–1037. doi: 10.1177/154193120504901107

Brysbaert, M., & Rastle, K. (2009). *Historical and conceptual issues in psychology*. Harlow, UK: Pearson/Prentice Hall.

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58*, 183–204. doi: 10.1901/jeab.1992.58-183

Chiesa, M. (1994). *Radical behaviorism: The philosophy and the science*. Sarasota, FL: Authors Cooperative.

Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language, 35*, 26–58. Retrieved from http://chomsky.info/articles/1967----.htm.

Cofer, C. N., & Foley Jr, J. P. (1942). Mediated generalization and the interpretation of verbal behavior. *The Psychological Review, 49*, 513–540. doi: 10.1037/h0059250

Darwin, C. (2006). *On the origin of species: By means of natural selection*. Mineola, NY: Dover Publications. (Original work published 1859)

Dinsmoor, J. A. (1985). The role of observing and attention in establishing stimulus control. *Journal of the Experimental Analysis of Behavior, 439*, 365–381. doi: 10.1901/jeab.1985.43-365

Dodge, R. (1907). An experimental study of visual fixation. *The Psychological Review Monograph Supplements, 8*, 1–96. doi: 10.1037/h0093042

Dodge, R., & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review, 8*, 145–157. doi: 10.1037/h0076100

Donahoe, J. W., & Palmer, D. C. (2004). *Learning and complex behavior*. Richmond, MA: Ledgetop Publishing.

Dougher, M., Twohig, M. P., & Madden, G. J. (2014). Editorial: Basic and translational research on stimulus–stimulus relations. *Journal of the Experimental Analysis of Behavior, 101*, 1–9. doi: 10.1002/jeab.69

Dube, W. V., Dickson, C. A., Balsamo, L. M., O'Donnell, K. L., Tomanari, G. Y., Farren, K. M. et al. (2010). Observing behavior and atypically restricted stimulus control. *Journal of Experimental Psychology, 94*, 297–313. doi: 10.1901/jeab.2010.94-297

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D. S., Balsamo, L. M., & Fowler, T. R. (1999). Eye tracking assessment of stimulus overselectivity in individuals with mental retardation. *Experimental Analysis of Human Behavior Bulletin, 17*, 8–14. Retrieved from http://www.eahb.org/bulletin/bulletin-archives/.

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D. S., Balsamo, L. M., Fowler, T. R. et al. (2003). Stimulus overselectivity and observing behavior in individuals with mental retardation. In S. Soraci & K. Murata-Soraci (Eds.), *Visual information processing* (pp. 109–123). Westport, CT: Praeger.

Dube, W. V., M., B. L., Fowler, T. R., Dickson, C. A., Lombard, K. M., & Tomanari, G. Y. (2006). Observing behavior topography in delayed matching to multiple samples. *The Psychological Record, 56*, 233–244. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*: Springer.

Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8–12. Retrieved from http://www.eahb.org/bulletin/bulletin-archives/.

Ferguson, B., Graf, E., & Waxman, S. R. (2014). Infants use known verbs to learn novel nouns: Evidence from 15- and 19-month-olds. *Cognition, 131*, 139–146. doi: 10.1016/j.cognition.2013.12.014

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*, 317–332. doi: 10.1901/jeab.1987.48-317.

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). New York, NY: Springer.

Guinther, P. M., & Dougher, M. J. (2015). The clinical relevance of stimulus equivalence and relational frame theory in influencing the behavior of verbally competent adults. *Current Opinion in Psychology, 2*, 21–25. doi: 10.1016/j.copsyc.2015.01.015

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-skinnerian account of human language and cognition*. New York, NY: Kluwer Academic/Plenum Publishers.

Hayhoe, M. M. (2004). Advances in relating eye movements and cognition. *Infancy, 6*, 267–274. doi: 10.1207/s15327078in0602_7

Holland, J. G. (1957). Technique for behavioral analysis of human observing. *Science, 125*, 348–350. doi: 10.1126/science.125.3243.348

Holland, J. G. (1958). Human vigilance. *Science, 128*, 61–67. doi: 10.1126/science.128.3315.61

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. New York: Oxford University Press.

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*, 107–124. doi: 10.1901/jeab.2006.58-04

Johnson, C., & Sidman, M. (1993). Conditional discrimination and equivalence relations: Control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 59*, 333–347. doi: 10.1901/jeab.1993.59-333

Kang, H. (2014). Understanding online reading through the eyes of first and second language readers: An exploratory study. *Computers & Education, 73*, 1–8. doi: 10.1016/j.compedu.2013.12.005

Keller, F. S., & Schoenfeld, W. N. (2011). *Principles of psychology: A systematic text in the science of behavior*. Cambridge, MA: CreateSpace (Original work published 1950)

Keller, M. (1943). Mediated generalization: The generalization of a conditioned galvanic skin response established to a pictured object. *The American Journal of Psychology, 56*, 438–448. doi: 10.2307/1417389

Kirshner, H., & Sidman, M. (1972). Scanning patterns in aphasic patients during matching-to-sample tests. *Neuropsychologia, 10*, 179–184. doi: 10.1016/0028-3932(72)90057-7

Madelain, L., Paeye, C., & Darcheville, J.-C. (2011). Operant control of human eye movements. *Behavioural Processes, 87*, 142–148. doi: 10.1016/j.beproc.2011.02.009

Majaranta, P., & Bulling, A. (2014). Eye tracking and eye-based human–computer interaction. In S. H. Fairclough & K. Gilleade (Eds.), *Advances in physiological computing* (pp. 39–65). London: Springer-Verlag

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience, 5*, 229–240. doi: 10.1038/nrn1348

Noton, D., & Stark, L. (1971a). Scanpaths in eye movements during pattern perception. *Science, 171*, 308–311. doi: 10.1126/science.171.3968.308

Noton, D., & Stark, L. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11*, 929–942. doi: 10.1016/0042-6989(71)90213-6

Palmer, D. C. (2010). Behavior under the microscope: Increasing the resolution of our experimental procedures. *Behavior Analyst, 33*, 37–45.

Perez, W. F., Endemann, P., Pessôa, C. V. B. B., & Tomanari, G. Y. (2014). Assessing stimulus control in a discrimination task with compound stimuli: Evaluating testing procedures and tracking eye fixations. *The Psychological Record, 65*, 83–88. doi: 10.1007/s40732-014-0092-1

Pessôa, C. V. B. B., Huziwara, E. M., Perez, W. F., Endemann, P., & Tomanari, G. Y. (2009). Eye-fixation to figures in a four-choice situation with luminance balanced area. *Journal of Eye Movement Research, 2*, 1–6. Retrieved from http://www.jemr.org.

Peters, H. N. (1935). Mediate association. *Journal of Experimental Psychology, 18*, 20–48. doi: 10.1037/h0057482

Pilgrim, C., & Galizio, M. (2000). Stimulus equivalence and unit of analysis. In J. C. Leslie & D. E. Blackman (Eds.), *Experimental and applied analysis of human behavior* (pp. 111–126). Reno, NV: Context Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422. doi: 10.1037/0033-2909.124.3.372

Reynolds, G. S. (1961). Attention in the pigeon. *Journal of the Experimental Analysis of Behavior, 4*, 203–208. doi: 10.1901/jeab.1961.4-203

Richardson, D. C., & Spivey, M. J. (2004a). Eye tracking: Characteristics and methods *Encyclopedia of Biomaterials and Biomedical Engineering* (Vol. 1, pp. 568–572). New York: Taylor & Francis.

Richardson, D. C., & Spivey, M. J. (2004b). Eye tracking: Research areas and applications *Encyclopedia of Biomaterials and Biomedical Engineering* (Vol. 1, pp. 573–582). New York: Taylor & Francis.

Riess, B. F. (1940). Semantic conditioning involving the galvanic skin reflex. *Journal of Experimental Psychology, 26*, 238–240. doi: 10.1037/h0063000

Riess, B. F. (1946). Genetic changes in semantic conditioning. *Journal of Experimental Psychology, 36*, 143–152. doi: 10.1037/h0060465

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols*. Paper presented at the Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, Palm Beach Gardens, Florida, USA.

Saunders, K. J. (1989). Naming in conditional discrimination and stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 51*, 379–384. doi: 10.1901/jeab.1989.51-379

Schroeder, S. R. (1970). Selective eye movements to simultaneously presented stimuli during discrimination. *Perception & Psychophysics, 7*, 121–124. doi: 10.3758/BF03210147

Schroeder, S. R., & Holland, J. G. (1968a). Operant control of eye movements. *Journal of Applied Behavior Analysis, 1*, 161–166. doi: 10.1901/jaba.1968.1-161

Schroeder, S. R., & Holland, J. G. (1968b). Operant control of eye movements during human vigilance. *Science, 161*, 292–293. doi: 10.1126/science.161.3838.292

Schroeder, S. R., & Holland, J. G. (1969). Reinforcement of eye movement with concurrent schedules. *Journal of the Experimental Analysis of Behavior, 12*, 897–903. doi: 10.1901/jeab.1969.12-897

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, and Hearing Research, 14*, 5–13. doi: 10.1044/jshr.1401.05

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis, 22*, 1–18. Retrieved from http://www.equivalence.net/pdf/Sidman_1987.pdf.

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15–27). Reno, NV: Context Press.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Authors Cooperative.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior, 37*, 23–44. doi: 10.1901/jeab.1982.37-23

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5–22. doi: 10.1901/jeab.1982.37-5

Skinner, B. F. (1981). Selection by consequences. *Science, 213*, 501–504. doi: 10.1126/science.7244649

Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*, 643–659. doi: 10.1901/jeab.1996.65-643

Tomanari, G. Y., Balsamo, L. M., Fowler, T. R., Lombard, K. M., Farren, K. M., & Dube, W. V. (2007). Manual and ocular observing behavior in human subjects. *European Journal of Behavior Analysis, 8*, 29–40. Retrieved from http://www.ejoba.org.

Velazquez, C. E., & Pasch, K. E. (2014). Attention to food and beverage advertisements as measured by eye-tracking technology and the food preferences and choices of youth. *Journal of the Academy of Nutrition and Dietetics, 114*, 578–582. doi: 10.1016/j.jand.2013.09.030

Wade, N. J. (2007). Scanning the seen: Vision and the origins of eye movement research. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye movements: a window on mind and brain* (pp. 31–61). Oxford: Elsevier Science.

Wade, N. J. (2009). Did Javal measure eye movements during reading? *Journal of Eye Movement Research, 2*, 1–7. Retrieved from http://www.jemr.org.

Wade, N. J. (2010). Pioneers of eye movement research. *i-Perception, 1*, 33–68. doi: 10.1068/i0389

Wade, N. J., & Tatler, B. W. (2005). *The moving tablet of the eye: The origins of modern eye movement research.* Oxford, NY: Oxford University Press.

Wells, W. C. (1792). *An essay upon single vision with two eyes: Together with experiments and observations on several other subjects in optics.* London, Cadell.

Wyckoff, L. B. (1952). The role of observing responses in discrimination learning. Part I. *Psychological Review, 59*, 431–442. doi: 10.1037/h0053932

Wyckoff, L. B. (1959). Toward a quantitative theory of secondary reinforcement. *Psychological Review, 66*, 68–78. doi: 10.1037/h0046882

Yarbus, A. L. (1967). *Eye movements and vision* (B. Haigh, Trans.). New York, N.Y.: Plenum Press.

Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation, 7*, 397–429. doi: 10.3758/BF03201553

Figure 1. Illustration of an afterimage. After fixating on the right dot placed upon the face of Robert Waring Darwin (1766–1848) fixating on the left dot will produce an afterimage. Picture retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3563053/figure/F20/ ©Wade, N. J., 2010, Creative Commons License.

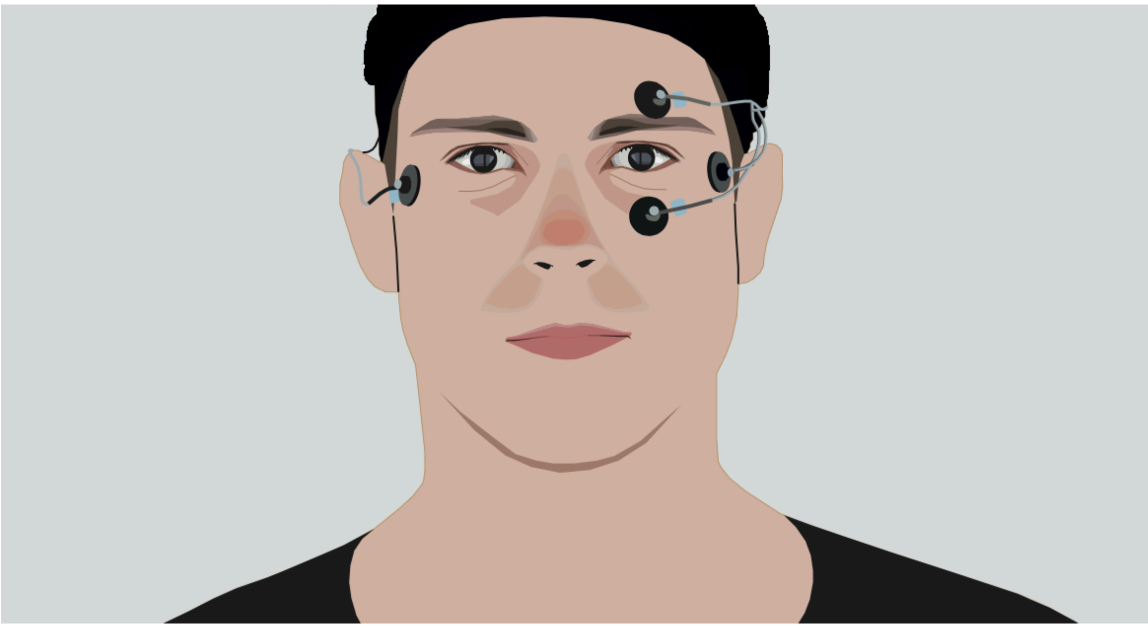

Figure 2. Illustration of a head mounted eye-tracker (electrooculography). Picture curtsey of Bjørving, M., 2015, adaptation of original picture retrieved from http://www.wired.co.uk/magazine/archive/2011/12/play/the-eyes-have-it/viewgallery/272206 ©Black Box, 2011.

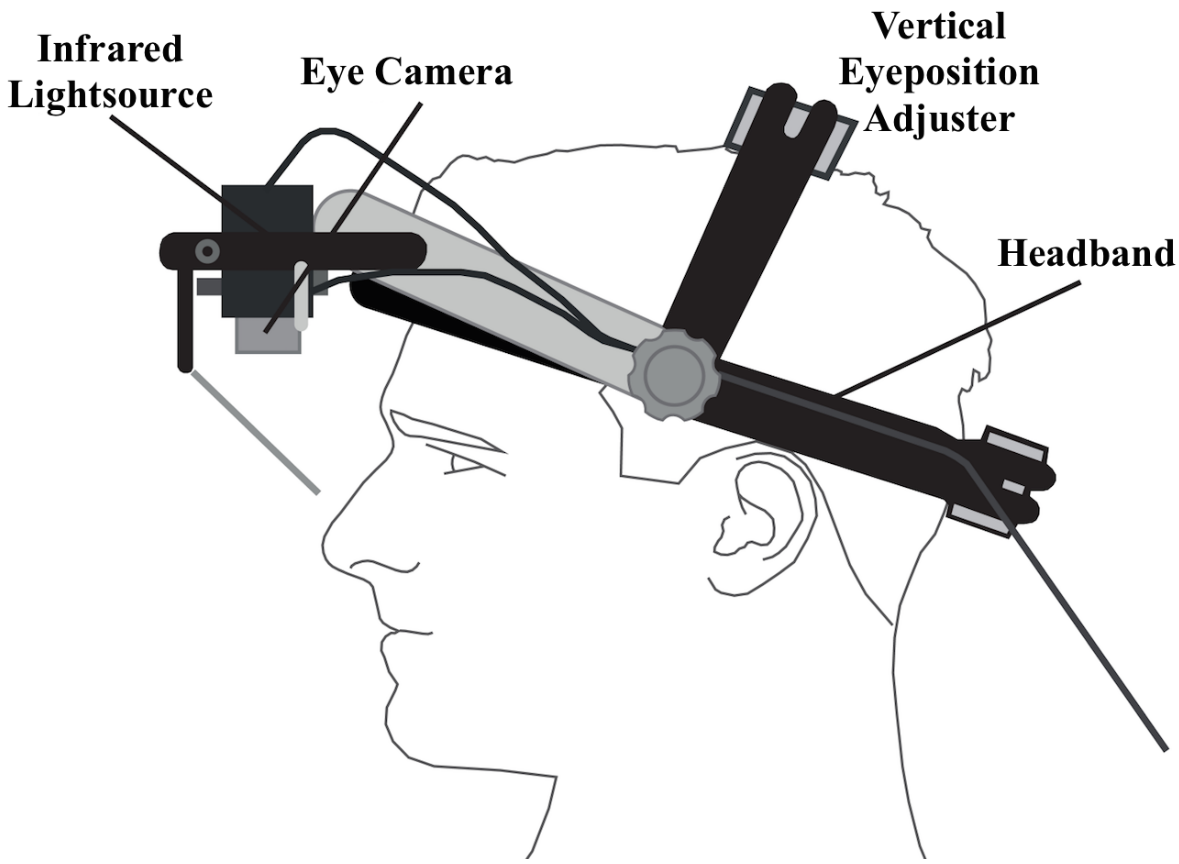

Figure 3. Illustration of a head mounted eye-tracker (photo-oculography/video-oculography) based on infrared technology. Picture curtsey of Bjørving, M., 2015, adaptation of picture retrieved from http://www.wikid.eu/images/d/d8/Eye_tracking_features.gif.

Complex Human Behavior:

Stimulus Equivalence and Eye-Tracking

The Effect of Training Structures on

Stimulus Equivalence and Ocular Observing Response

Pedram Sadeghi

Oslo and Akershus University College

Abstract

Research pertaining stimulus equivalence and the effect of training structures on the establishment of equivalence class formation has shown that there is minimal differences between the many-to-one (MTO) and the one-to-many (OTM) training structures, while the linear series (LS) has shown to be the least efficient training structure in producing equivalence class formation. Moreover, some research has shown that an observing response can affect how fast stimulus equivalence is established. The current study utilized eye-tracking—a seldom used technology within behavior analysis to expand the analysis of observing responses—with the aim of investigating how different training structures affect the establishment of equivalence class formation and ocular observing response. Thirty participants were randomly assigned to undergo one of the three training structures with the aim of establishing three 5-member classes. The results were consistent with previous research finding, as 7 out 10 participants in the MTO group and 10 of 10 participants in the OTM group were successful in responding in accordance with stimulus equivalence, while none of the participants in the LS group showed responding in accordance with stimulus equivalence. Furthermore, eye-tracking data showed a stepwise reduction of ocular observing responses during training, which is typically observed with regard to reaction time. Eye-tracking data also gave a more detailed analysis of how the different groups performance of ocular observing responses came about during the testing of stimulus equivalence.

Keywords: Stimulus equivalence, training structure, eye-tracking, ocular observing response, linear series, many-to-one, one-to-many, attention, stimulus control.

The effect of training structures on stimulus equivalence and ocular observing response

Stimulus equivalence is defined by responding with the properties of reflexivity, symmetry, and transitivity (Sidman & Tailby, 1982). Until now there has accumulated a torrent of findings, displaying the numerous inconspicuously subtle variables that have an effect on stimulus equivalence formation (Arntzen, 2012).

Although Sidman and Tailby (1982) recognized that class size would have an effect on establishing stimulus equivalence, they did not consider training structures to have an effect on stimulus equivalence outcomes. The three most common training structures for establishing the prerequisites, i.e., conditional discriminations, for stimulus equivalence are the linear series (LS); the many-to-one (MTO); and the one-to-many (OTM) training structures. In the LS training structure one sample stimulus is trained to a comparison stimulus, this comparison stimulus is later presented as a sample stimulus when presenting a third stimulus as a comparison stimulus. The stimulus presented as both sample stimulus and comparison stimulus is called a node since it is linked to two other stimuli (Fields & Verhave, 1987). With the MTO training structure, many samples stimuli are trained to a fixed comparison stimulus. In the OTM training structure it is the opposite; one fixed sample stimulus is trained to multiple comparison stimuli.

R. R. Saunders and Green (1999) *Discrimination Analysis* provided an elegantly parsimonious account for why different training structures could give different yields on tests for stimulus equivalence. This account was based upon numbers of simultaneous simple discriminations and successive simple discriminations. The discrimination analysis holds that each training structure constitutes different numbers of simultaneous simple discriminations and successive simple discriminations. While the MTO training structure includes training of all the simple discriminations, the LS and OTM training structure do not. Since the numbers of relations are exponentially expanded, these differences will be more prominent when there are larger numbers of classes or members in a class. To illustrate this exponential difference one can consider the establishment of three 3-member classes, which requires 36 discriminations. In the LS and OTM training structures, 27 of 36 discriminations will be trained (i.e., nine discriminations are not presented during training); while all the discriminations are trained with the MTO training structure. On the other hand when establishing three 5-member classes—as is the case with the current study—with the LS and OTM training structures only 51 out of 105 discriminations will be trained (i.e., 54 discriminations are not presented); conversely for the MTO training structure all the discriminations are presented. Furthermore there are four times as many simple simultaneous discriminations of comparison stimuli in the OTM training structure compared to the MTO training structure, likewise the opposite is observed when considering successive discriminations of sample stimuli, where there are four times as many successive discriminations for the MTO training structure compared to the OTM training structure. On the basis of the aforementioned—i.e., that each training structure determines how many numbers of simultaneous and successive discrimination are going to be trained—the MTO training structure should be the most efficient, and differences between the MTO and the OTM training structure would be more salient as number of classes and/or class size would become larger.

Still there have been inconsistencies pertaining to the role of training structures in establishing stimulus equivalence. Some have found that the MTO training structure is the most efficient (Fields, Hobbie-Reeve, Adams, & Reeve, 1999;

Hove, 2003; K. J. Saunders, Saunders, Williams, & Spradlin, 1993; R.R. Saunders, Chaney, & Marquis, 2005; R. R. Saunders, Drake, & Spradlin, 1999; R. R. Saunders & Green, 1999; R. R. Saunders & McEntee, 2004; Spradlin & Saunders, 1986); while some results showed that OTM training structure is the most efficient (Arntzen & Holth, 1997, 2000); and others that there are minimal differences between the MTO and the OTM training structures (Arntzen, Grondahl, & Eilifsen, 2010; Arntzen & Hansen, 2011; Arntzen & Nikolaisen, 2011; Arntzen & Vaidya, 2008; Smeets & Barnes-Holmes, 2005). Altogether the LS training structure has shown to be the least efficient in establishing stimulus equivalence (Arntzen, 2012; Arntzen et al., 2010; Arntzen & Holth, 1997, 2000). It must be noted that one of the procedural variables that are in interplay, which could affect the outcomes on test for stimulus equivalence are the number of possible choices, i.e., numbers of comparison stimuli. The studies where the MTO training structure has shown to be superior, have predominantly utilized what is called two-choice MTS. Numbers of comparison stimuli is an important factor to take into consideration, since two-choice MTS increases the likelihood of a responding correct by chance, furthermore it could produce rejection control (Carrigan & Sidman, 1992; Johnson & Sidman, 1993).

It is of equal importance to note that these different results only are observed when a simultaneous training and testing protocol is employed—another variable that affects stimulus equivalence formation are the three training and testing protocols, in which establishment of conditional discriminations can be established and the formation of stimulus equivalence classes can be tested—i.e., simple-to-complex; the complex-to-simple; and the simultaneous protocol. In the simple-to-complex protocol, each individual relation is introduced trained then tested sequentially in an incremental order, testing for symmetry, transitivity, and equivalence, before yet another new relation is introduced. In contrast complex-to-simple protocol it is reversed, it starts with an equivalence test, which if is not passed results in testing of symmetry and transitivity. The protocols are concluded with a mixed test. Based upon results from the tests it is determined if stimulus equivalence class has been established or not. In the simultaneous protocol, all baseline relations are trained simultaneously before symmetry, transitivity, and equivalence relations is tested in a mixed block (Imam, 2006). It is the latter of the three protocols, which has been implemented most frequently in the stimulus equivalence paradigm (Arntzen, 2010).

The understanding of these variables, specifically training structures is of critical importance for both the understanding of stimulus equivalence in itself, as well as for the enhancement of applied applications (Guinther & Dougher, 2015). Especially concerning applied programs where an increase of stimuli results in an exponential expansion of not directly trained relations, which are of immense value.

Dymond and Rehfeldt (2001) proposed that employing novel additional measures (e.g., reaction time) could advance and enhance the understanding of stimulus equivalence, and that relying on only one measure possibly could hinder new discoveries pertaining to the nature of stimulus equivalence. Furthermore, they stated that additional measures could be have beneficial value for those that develop applied programs based on stimulus equivalence. The use of measurements like reaction time, which is not a part of the stimulus equivalence definition, have given broader insight into the different relations and even to how stimulus equivalence comes about and is maintained (e.g., Arntzen, Petursson, Sadeghi, & Eilifsen, 2015; Dymond & Rehfeldt, 2001; Eilifsen & Arntzen, 2009; Pilgrim & Galizio, 2000; Spencer & Chase, 1996). For instance Spencer and Chase (1996) assessed the reaction time of responding and found an increment in the reaction time to comparison stimuli, from the symmetry trials to the equivalence trials, which have been reconfirmed by later research (Arntzen, Braaten, Lian, & Eilifsen, 2011; Arntzen et al., 2010; Arntzen & Hansen, 2011; Arntzen & Holth, 1997, 2000; Eilifsen & Arntzen, 2009; Holth & Arntzen, 1998, 2000; R.R. Saunders et al., 2005; Spencer & Chase, 1996). Thereby as a result of taking reaction time data into considerations, there is ample evidence indicating that different relations have different intrinsic properties (Pilgrim & Galizio, 2000).

An additional measure that seldom has been utilized in behavior analysis is eye-tracking (Palmer, 2010). Eye-tracking is a measuring procedure that identifies where and how a person is seeing within a visual scene, which has been adopted as a measuring instrument and frequently employed by a number of scientific disciplines such as cognitive psychology, neuroscience, and infant research. The advantage with eye-tracking is that it gives a multifaceted access to measurements of moment-to-moment behaviors which quintessentially are salient in “day to day life”. Furthermore, eye-tracking allows for stronger inferences than measurement of percentage of correct responding or that of reaction time, since it grants access to a more detailed assessment of how the behavior occurs (Hayhoe, 2004; Kirshner & Sidman, 1972).

On the occasions eye-tracking has been employed it has been for assessing stimulus control relations, often in close relation to the concept of the observing response (Dube et al., 2010; Dube et al., 1999; Dube et al., 2003; Dube et al., 2006; Kirshner & Sidman, 1972; Perez, Endemann, Pessôa, & Tomanari, 2014; Pessôa, Huziwara, Perez, Endemann, & Tomanari, 2009; Schroeder, 1970; Schroeder & Holland, 1968a, 1968b, 1969; Tomanari et al., 2007). In the stimulus equivalence paradigm, an observing response could be a requirement of pressing the sample stimulus so that comparison stimuli are presented on a display. The observing response is not reinforced directly; its function is to give access to discriminative stimuli. For example in a MTS situation the correct comparison stimulus is the Sd for correct responding, while at the same time it is a secondary reinforcing stimulus for pressing the sample stimulus (Dinsmoor, 1983, 1985; Dinsmoor, Bowe, Green, & Hanson, 1988; Holland, 1957; Wyckoff, 1952, 1959). According to Dinsmoor (1985) observing response is a requirement for the establishment of accurate visual discrimination. Similarly, recent findings within stimulus equivalence have shown that a requirement of observing responses can affect how quickly the conditional discriminations are learned (Arntzen et al., 2011). With the use of eye-tracking, one can expand the observing response to an ocular observing response.

Moreover, seminal research within eye-tracking has shown that tasks participants are exposed to, impact their eye movements (Buswell, 1935; Yarbus, 1967). Similarly, Sidman (1994) remarked that for participants with intellectual disabilities and children of the age of six and younger, errors often occurred since the participants failed to observe all the relevant stimuli—in which made him and colleagues (e.g., Kirshner & Sidman, 1972) curious to whether their inadequacies in emitting the necessary ocular observing response, which they referred to as scanning, where caused by lack of learning experience or damage to the central nervous system.

In a recent study Steingrimsdottir and Arntzen (accepted) incorporated eye-tracking to assess differences between how younger and older adults´ ocular observing response come about, i.e., fixation duration and fixation rate, when training and testing for stimulus equivalence class formation. They used the MTO training structure to train five 3-member classes. Their results showed a decline of ocular observing responses across the training; during the testing phase a stepwise increment was observed both in start of the test for the baseline, symmetry, and equivalence relations; the same increase was observed at the end of the test, the difference being that there was a decrement of ocular observing responses at the end of the test relative to beginning.

The purpose of the current study is to assess how different training structures can affect stimulus equivalence formation, as well as expanding the depth of analysis with the implementation of eye-tracking to assess how different training structures affect ocular observing response. As far as the author is aware, this is the first experiment within the stimulus equivalence paradigm that has utilized methods of eye-tracking to specifically study the effect of training structures on stimulus equivalence formation and ocular observing response. One prediction is that the different training structures call for different types of ocular observing response, i.e., affects how participants attend to the stimuli. The current experiment is important for understanding of the seemingly elusory role of training structures, which is critical both conceptual and application wise. Also, in addition to using novel experimental procedure, i.e., eye-tracking, within the stimulus equivalence paradigm—which previously has shown to be viable means of assessing stimulus control relations—the current study expands the analysis with regard to eye movements called transitions, which earlier has been referred to as a scanning behavior.

Method

Participants

Thirty[1] participants were recruited, these were students recruited on campus, and through personal contacts, before being randomly assigned to one of three groups.

[1] Overall 53 participates were recruited, of these, 13008, 13010, 130014, 13021, 13023, 13026, 13027, 13029, 13037, 13042, 13044, 13055, were excluded due to data loss in relation with eye-tracking; 13004, 13007, and 13012 because of software malfunction; and 13030, 13032, 13048, and 13049 wished to cancel.

The participants 21 males and nine females ranging between 19–29 years of age, with an average age of 24 years. Participants´ average age in the LS group (i.e., participants 13003, 13011, 13016, 13020, 13024, 13028, 13033, 13038, 13043, and 13051) was 23.5 years; MTO group (i.e., participants 13003, 13011, 13016, 13020, 13024, 13028, 13033, 13038, 13043, and 13051) was 23.5 years; and 24.5 years for OTM group (i.e., participants 13005, 13006, 13009, 13018, 13022, 13034, 13035, 13045, 13047, and 13053), respectively. They were recruited on the basis of their willingness of participating and lack of knowledge of stimulus equivalence[2]. Upon recruitment participants were told that the purpose of the experiment was to see how learning comes about, that they would be required to sit in front of a computer and wear eye-tracking glasses, that their identity would be kept completely anonymous, that they would be able to resign from the experiment at any moment of their choosing, and that at the end of the experiment they were to receive a debriefing where the purpose of the experiment would be explained.

Setting

The experiment was conducted in a soundproof room with the size of 4.6 m^2 (1.02 m x 4.48 m), where all superfluous stimuli were removed, and thick dark drapes covered the two windows in the room, to hinder sunlight and thereby ensuring stable illumination throughout the entire experiment. In front of where the participants would sit, there was placed a chin rest with a distance of 60 cm from the display.

Design

[2] Participants 13001–13003 and 13020 had been in similar experiments, i.e., they were familiar with the MTS format but did not possess knowledge about stimulus equivalence.

The experimental design was a between-group design, consisting of three groups with 10 participants, where each group defined the experimental condition (i.e., training structure of LS, MTO, and OTM) they would undergo when establishing baseline relations (BSL) (i.e., directly trained relations).

Apparatus and Software

The experiment was conducted on a custom built PC with AMD® Athlon™ II X2 250 3.0 GHz CPU and 4 GB RAM, running Windows 7 Professional (32-bit), which was connected via AVerKey 300, to an ISCAN® computer with Intel® Pentium® 4 3.4 GHz CPU and 1 GB RAM, running Windows 7 Professional (32-bit). Where the former was running MTS program provided with curtsey of Professor Erik Arntzen (Oslo and Akershus University College, Norway), and the eye-tracking analyzer program, the latter was running an ISCAN® DQW version 1.2 which was used for calibrating the eye-tracking glasses and process eye-tracking data. The eye-tracker was an ISCAN® head-mounted pupil/corneal reflection eye-tracking system (ISCAN Corp., Burlington, MA; http://www.iscaninc.com).

Stimulus Material

The 15 visual stimuli consisted of letters from the Greek, Japanese, Arabic, and Cyrillic alphabets that potentially could form three 5-member classes (see Figure 1). Stimuli were presented on 17 inch LCD with screen resolution of 1280 x 1024, with height and width of ca. 2.5 x 2.5 cm, and a distance of ca. 16.5 cm from outer to outer edge of the stimuli. Each stimulus was presented in an area with height and width of 5.6 x 4.1 cm, which amounted the Area of Interest (AOI).

Procedure

General information to participants.

Prior to conducting each session, participants were asked to read an information pamphlet, informing them of their rights, the format for the experiment, and a consent form that they had to sign. Participants were informed that the experiment was in conjunction with a master thesis concerning the psychology of learning, with the purpose of on examining the variables that affect stimulus equivalence. They were told that the length of the experiment was about 1.5 to 2.5 hours, depending on which group they were assigned to. Moreover, they were told that they would be anonymous and that they could withdraw from the experiment at any time. In order not to affect the experiment, stimulus equivalence was not explained any further, but the participants were told that they would receive a thorough debriefing of results and explanation of stimulus equivalence, upon which they also could ask questions. The debriefing included the history, core concepts, and the applied application of stimulus equivalence, as well as offering them a book chapter or an article explaining stimulus equivalence in further detail.

Calibration.

After the briefing, each participant was equipped with the eye tracker and asked to sit in front of the computer, and to rest their head on the chin rest. The eye-tracking glasses were calibrated with the use of a picture with five points displayed on the screen. Participants were instructed to fixate on each specific dot while the experimenter observed the gaze of the participant.

Instructions.

When calibration was successful the participants were told to press the “Start” button when they were ready, whereupon the MTS-program would start by presenting this instruction:

"A stimulus will appear in the middle of the screen. Click on this by using the computer mouse. Three other stimuli will then appear. Choose one of these by using the computer mouse. If you choose the stimulus that we have defined as correct, words like *good*, *super*, etc. will appear on the screen. If you press a wrong stimulus, the word *wrong* will appear on the screen. During some stages of the experiment, the computer will not tell you if your choices are correct or wrong. However, based on what you have learned, you can get all the tasks correct. Do the best you can to get the most correct. Good Luck!"

Training and testing.

With the utilization of a simultaneous protocol, each trial began with a presentation of a sample stimulus in the middle of the display, which upon pressing three comparison stimuli would appear randomly in the corners of the display. After responding, feedback of programmed consequences would be given for 500 ms. The inter-trial interval (ITI) was 500 ms both during the training phase and the testing phase. All BSL relations were presented three times resulting in 36 trials in each block. In the training phase, the mastery criterion was 33 correct responses out of a block of 36 trials (i.e., 91%). Trials in the training phase were presented in a concurrent trial arrangement (i.e., all BSL relations were trained at the same time). With the onset of each trial, the mouse cursor would be repositioned back to the middle of the display.

When reaching 100% correct responding fading would commence, first with only 75% programmed consequences, then 50% programmed consequences, and then a block with 0% programmed consequences. Upon completion of the training phase, the test phase would begin. The test phase consisted of 180 trials which of 36 were BSL trials; 36 were symmetry (SYM) trials; 108 were equivalence formation (EQF) trials for the MTO and OTM group; and 54 transitivity (TRA) trials and 54 equivalence (EQ) trials for the LS group. No programmed consequences were given during the test phase. The criterions for responding in accordance with the various relations in the test were 98 out of 108 EQF trials; 33 out of 36 for both BSL and SYM trials respectively; and 49 out of 54 for TRA and EQ, respectively for the LS group.

Data analysis.

During the experiment data collected included reaction time (RT), which was provided by the MTS software; also fixation time (FT); and fixation rate (FR), were acquired using identification by dispersion threshold algorithm (I-DT), with a threshold of 200 ms[3]; furthermore transition rate [4] (TR) which was acquired using visual inspection with an eye-tracking analyzer. RT was the time from the participants responded to the sample stimulus, to responding to one of the comparison stimuli. FT was the total duration the participants gazed at each AOI during a single trial. In the test phase, FT data was based upon first five and last five trials with fixation time over

[3] Due to drifts and fixation in vicinity of AOI some data were acquired using visual inspection for some participants (i.e., 13017, 13019, 13024, 13033, 13034, 13035, 13038, 13039, 13043, 13051, and 13053). The IOA for visual inspection was 91%. Due to data loss of FT data for the penultimate block was added for 13017. According to Salvucci and Goldberg (2000) the fixation duration threshold values of 100–200 ms is consistently reported in Eye-Tracking research.

[4] According to Holmqvist and Nystrøm (2011) a transition, (also called a gaze shift), is the movement from one AOI to another AOI. One trial in the first training block was omitted because of data loss for participants 13034 and 13038.

the 200 ms threshold. FR was the number of times participants gazed at each AOI during a single trial. TR was the number of times participants looked between the AOIs. Mean values for each groups RT and FT data were based upon the median of the 36 trials in the first block; 36 trials in the first block where the participants responded above 60% correct (i.e., DG>0.60 block[5]); 36 trials in the last training block; in addition the first five -BSL trials; -SYM trials; and -EQF trials; and last five -BSL trials; -SYM trials; and -EQF trials in the test phase, for each participant. Group mean values of FR and TR data were calculated by taking the total number of eye-movements and dividing them by number of trials (e.g., 36 for entire blocks; and 5 for specific trial types), for each participant. The reason for including the DG>0.60 block, which was an arbitrarily chosen number the lab operated with, was to monitor changes in the experiment.

Results

Number of Training Trials and Errors

The average number of trials needed for participants in each group to reach the test phase was, 587 for the LS group; 652 for the MTO group; and 482 trials for the OTM group, respectively (see Table 1–3).

During the training phase the average number of errors for participants from each group were 182 for the LS group; 207 for the MTO group; and 170 errors for the OTM group, respectively.

Independent samples *t*-test on numbers of training trials and errors showed no differences as a function of training structure.

Derived relations

[5] DG = discrimination gradient.

Results from the test phase showed that for the LS group 5 of 10 participants the BSL relations were still intact, 1 out of 10 participants that responded in accordance with SYM, and 0 out of 10 in accordance with neither TRA nor EQ (see Table 1).

The MTO group had 8 out of 10 participants responding showed that the BSL relations were still intact, 9 out of 10 participants that responded in accordance with symmetry, and 7 out of 10 in accordance with EQF (see Table 2).

All the participants in the OTM group showed that the BSL relations were still intact, in addition all the participants responded in accordance to both SYM and EQF (see Table 3).

The results show a significant relationship between different training structures and yields on equivalence formation, X^2 (1, N=30) = 21.448, $p < .05$, Cramer's V = .84. Furthermore, a Chi test for BSL was 0,029002517; for SYM was 0,000017558; and EQF was 0,000022011, respectively, showed a strong correlation between training structure and high yields in SYM and EQF trials.

Reaction Time

RT data showed that for the first block participants from the LS group used 4402 ms; MTO group used 4712 ms; and OTM group used 4137 ms, respectively (see Figure 2). On the DG>0.60 block, participants in the LS group used 3534 ms; the MTO group used 2199 ms; and the OTM group used 4003 ms. On the last block the participants from the LS group used 1947 ms; MTO group used 1422 ms; and OTM group used 2204 ms, respectively.

RT of responding during the test showed that for the first 5 BSL trials, the participants in the LS group used 3142 ms; the MTO group used 2681 ms; and the

OTM group used 3285 ms. On the last 5 BSL trials the LS group used 1999 ms; the MTO group used 2843 ms; and the OTM group used 2086 ms.

Further on, for the first 5 SYM trials the participants from the LS group used 5106 ms; MTO group used 4576 ms; and the OTM group used 3331 ms. On the last 5 SYM trials the participants in the LS group used 2341 ms; the MTO group used 2440 ms; and the OTM group used 2284 ms.

When responding on the first five EQF trials the RT for the participants from LS group was 6289 ms; the MTO group was 6193 ms; and the OTM group was 5725 ms. On the last 5 EQF trials the RT are, 2519 ms for the participants in the LS group; 4009 ms for the MTO group; and 2722 ms for the OTM group. In addition, the participants from the LS group responds faster on the five last TRA trials, i.e., 6580 ms, relative to the five first TRA trials 2592 ms.

Fixation Time

On the first block, FT was 715 ms for the participants in the LS group; 837 ms for the participants from the MTO group; and 629 ms for the participants from the OTM group (see Figure 2). The FT was 547 ms, on DG > 0.60 blocks, for the participants in the LS group; 496 ms for the participants in the MTO group; and 567 ms for the participants from the OTM group. In the last block FT was 390 ms for the participants from the LS group; 417 ms for the participants from the MTO group; and 482 ms for the participants in the OTM group.

In the test phase, the first five BSL trials the FT for the participants in the LS group was 623 ms; FT was 439 ms for the participants in the MTO group; and 496 ms for the participants from the OTM group. On the last five BSL trials in test phase the FT was 505 ms for the participants in the LS group; 444 ms for the participants from the MTO group; and 421 ms for the participants in the OTM group, respectively.

FT for the first five SYM trials for the participants from the LS group was 713 ms; 560 ms for the participants in the MTO group; and for the participants in the OTM group it was 552 ms, respectively. In the last five SYM trials FT was 422 ms for the participants from the LS group; for the participants from the MTO group it was 542 ms; and 461 ms for the participants in the OTM group.

The FT for the first five EQF trials for the participants in the LS group was 720 ms; 608 ms for the participants in the MTO group; and 546 ms for the participants from the OTM group, respectively. For the last five EQF trials the FT for the participants from the LS group were 487 ms; 533 ms for the participants in the MTO group; and 523 ms for the participants from the OTM group. In addition for the LS group, the FT on the first five and last five TRA trials was 619 ms; and 531 ms, respectively.

Fixation Rate

FR in the first block for participants from the LS group was 2.0; for participants from the MTO group the FR was 2.2; and 2.0 for participants in the OTM group, respectively (see Figure 2). In the DG>0.60 block, participants in the LS group the FR was 1.5; 0.9 for the participants in the MTO group; and 1.4 for the participants from the OTM group. In the last block, the FR was 0.7 for participants from the LS; 0.5 for participants from the MTO group; and participants in the OTM group had 0.8, respectively.

FR when responding on the first five BSL trials was 1.2 for participants from the LS group; 0.9 for participants from the MTO group; and 1.0 for participants in the OTM group, respectively. FR was 1.2 when responding on the last five BSL trials for participants from the LS group; 1.1 for participants in the MTO group; and 0.8 for the participants from the OTM group.

First five SYM trials FR was 1.9 for the participants from the LS group; 1.4 for the participants from the MTO group; and 1.0 for the participants in the OTM group. On the last five SYM trials participants in the LS group showed FR of 0.9; FR was 0.7 for participants from the MTO group; and 0.7 for participants in the LS group.

FR for the first five EQF trials for the participants in the LS group was 2.4; 2.0 for the participants from the MTO group; and 2.0 for the participants in the OTM group. On Last five EQF for the participants from the LS group FR was 0.9; while for the participants from the MTO group it was 1.5; and 0.9 for the participants in the OTM group. Furthermore for the LS group the five first and last TRA trials were 2.3; and 1.0 respectively.

Transition Rate

On the first block, the TR was 0.62 for the participants from the LS group; 0.68 for the participants in the MTO group; and 0.76 for the participants from the OTM group (see Figure 2). In the DG > 0.60 block for the participants in the LS group the TR was 0.40; 0.23 for the participants in the MTO group; and 0.46 for the participants in the OTM group. On the last block the participants in the LS group had a TR of 0.18; the participants from the MTO group the TR were 0.05; and 0.23 for the participants in the OTM group.

TR on the first five BSL trials for the participants from the LS group was 0.20; for the participants from the MTO group it was 0.14; and 0.44 for the participants from the OTM group. For the last five BSL trials, TR was 0.36 for the participants in the LS group; 0.14 for the participants from the MTO group; and 0.12 for the participants in the OTM group.

For the first five SYM trials TR for the participants from the LS group was 0.36; for the participants in the MTO group the TR was 0.46; and 0.26 for the participants from the OTM group. TR was 0.22 in the last five SYM trials for the participants in the LS group; 0.30 for the participants from the MTO group; and TR of 0.10 for the participants from the OTM group, respectively.

In the first five EQF trials for the participants in the LS group TR was 0.50; for the participants from the MTO group TR was 0.52; and 0.58 for the participants from the OTM group. TR was 0.36 for Last five EQF trials for the participants from the LS group; 0.54 for the participants from the MTO group; and 0.28 for the participants in the OTM group. On the first five TRA trials TR was 0.62: and 0.22 on the five 5 TRA trials, for the participants in the LS group.

Statistics of Additional Measures

Paired samples statistics *t*-test were conducted for RT (see Table 4); FT (see Table 5); FR (see Table 6); and TR (see Table 7) between the last training block, and the first five -BSL; -SYM; and –EQF (EQ) trials; as well as between the first five -BSL; -SYM; and -EQF trials and the last five -BSL; -SYM; and –EQF (EQ) trials, across all participants. The statistics showed significant differences across nearly all the additional measures, with the exception being: (1) FT, between the first five BSL trials and last five BSL trials; between the first five SYM trials and last five SYM trials, (2) FR, between the first five BSL trials and last five BSL trials; and (3) TR, between first five BSL trials and last five BSL trials; between the first five SYM trials and last five SYM trials; between the first five EQF trials and last five EQF trials, in the test phase, where there was no significant difference.

Discussion

The purpose of this study was to extend the analysis of how different training structures affect the establishment of stimulus equivalence, and to provide a more meticulous methodological expansion in the form of eye-tracking. The LS group was the group which participants showed highest percentage correct responding during training, and least efficient in producing responding in accordance with stimulus equivalence (i.e., 0 of 10 participants). Therefore the data supports previous findings of the LS training structure being the least efficient in establishing stimulus equivalence and the Saunders and Green´s discrimination analysis (i.e., Arntzen, 2012; Arntzen et al., 2010; Arntzen & Holth, 1997, 2000; R. R. Saunders & Green, 1999). Participants in the MTO group needed most training trials and had the highest numbers of errors out of the three groups. The fact that the MTO participants required most training trials and numbers of emitted errors are consistent with the discrimination analysis (R. R. Saunders & Green, 1999). Still, only slightly over half of the participants in the MTO group showed responding in accordance with stimulus equivalence (i.e., 7 out of 10 participants). An explanation of these findings may be as Arntzen and Vaidya (2008) have proposed; that the reason for high yields for OTM participants are due to the fact that adult participants usually have a learning history in which provides them with a rich behavioral repertoire. Furthermore, in accordance with the predictions of the discrimination analysis the OTM group the needed least amount of training trials, and showed the lowest amount of errors during training. Participants from the OTM group were the most efficient in responding in accordance with stimulus equivalence (i.e., 10 out of 10 participants), which reconfirm earlier research that has shown minimal differences between the MTO and the OTM training structures when considering yields on stimulus equivalence test (i.e., Arntzen et al., 2010; Arntzen & Hansen, 2011; Arntzen & Nikolaisen, 2011; Arntzen & Vaidya, 2008; Smeets & Barnes-Holmes, 2005), yet they neither support discrimination analysis, nor earlier research were the MTO training structure was superior in establishing stimulus equivalence class formation (i.e., Fields et al., 1999; Hove, 2003; K. J. Saunders et al., 1993; R.R. Saunders et al., 2005; R. R. Saunders et al., 1999; R. R. Saunders & Green, 1999; R. R. Saunders & McEntee, 2004; Spradlin & Saunders, 1986). In conclusion the results from the current study did indeed replicate an ongoing trend where the OTM training structure seems to be slightly more efficient in favorable outcomes in establishing stimulus equivalence.

Reaction Time

Even though the participants in the MTO group responded slowest at the start of the training phase they were fastest at the end, while the opposite is observed for the participants from the OTM group, which is in concordance with the discrimination analysis considering the numbers of discriminations being trained. Contrary to the participants from the LS and the OTM group, which responded over 1 s faster on the five last BSL trials relative to the five first BSL trials; the participants in the MTO group responds 162 ms slower.

All the participants responded faster on the five last SYM trials, relative to the five first SYM trials (i.e. LS 3 s; MTO 2 s; OTM 1 s, faster responding respectively). Also the participants in the LS group were the slowest of the three groups while the participants in the OTM group were the fastest.

The participants in the MTO group went from being the fastest group to being the slowest, regarding EQF trials, while the participants from the LS group went from being the slowest group to being the fastest group. When considering RT across groups the same pattern is shown, faster responding from the beginning to the end of the test; that during the test the RT on the BSL trials are longer relative to BSL trials at the end of the training; likewise SYM trials are of higher RT relative to the BSL trials; and moreover, that the RT in the EQF trials are higher RT relative to the SYM trials. In sum the study replicated previous findings on RT (Arntzen et al., 2010;

Arntzen & Hansen, 2011; Arntzen & Holth, 1997; Eilifsen & Arntzen, 2009; Holth & Arntzen, 1998, 2000; R.R. Saunders et al., 2005; Spencer & Chase, 1996)

Fixation Time

FT data were in line with the RT data, in the sense that even though the participants in the MTO group attended to the comparison stimuli for longest duration at the start of the training phase, they were the group to attended in shorter duration at the end of the training phase relative to the participants in the OTM; which showed more stable development over the course of the training. The distinct reduction of 50% from the first block to the last block for the participants in the MTO group may be explained by the how all the relations are trained in the MTO training structure compared to the other training structures. The variability in the training trials causing longer fixation times would be consistent with the fact that there are four times as many successive discriminations in the MTO training structure relative to that of the OTM training structure. The numbers of discriminations needed to be learned before test are introduced could also explain why participants in the OTM group had the least amount of change during training, with a reduction of 23% from the first to last training block.

Out of all three groups, the participants in the LS group attended with the least duration of time on each comparison stimuli at the end of the training phase. When assessing FT for the LS group in the training phase (the three temporal points), there seems to be a relatively fixed (roughly 162 ms) step size. The cause of this could be the fact that in the LS training structure, some stimuli are nodes, i.e., they function both as sample stimuli and as comparison stimuli (Fields & Verhave, 1987). Thereby when one relation is learned there is the possibility that other relations involving the same node do not require the same duration of ocular observing responses as that of the previous relation. In general the decline of FT throughout the training phase across the groups, are consistent with the findings of Steingrimsdottir and Arntzen (accepted).

When considering FT and the BSL trials the MTO distinguished itself, i.e., that the participants actually looked for a duration of 5 ms longer on the five last BSL trials relative to the five first BSL trials; while the LS: and OTM Showed opposite behavior. At the end of the training phase, participants in the OTM group were the group to show highest FT; while the MTO group showed the highest reduction of FT over the course of the training phase, and was the group with the lowest FT.

Regarding the SYM and EQ trials, participants in the LS group fixated on the comparison stimuli for the longest duration on the first five trials, but shortest duration when fixating on the last five trials. Furthermore, the participants from the MTO group fixated longest on last five trials of both SYM and EQF trials relative to the other two groups. When considering FT for BSL and the derived relations SYM and EQF across the three groups, the results replicate the findings of Steingrimsdottir and Arntzen (accepted), with the exception of the end of the test phase, where the LS groups FT was of longer duration for BSL trials relative to SYM trials. This could be the effect of training structures indirectly shaping some kind of eye movements, which carry over to the test phase.

Fixation Rate

While there were no differences in FR for the LS group i.e. 1.2; FR was 0.2 higher for the participants from the MTO group on the last five BSL trials, relative to the first five BSL trials; while the participants from the OTM group showed a 0.2 decrease in FR.

At the end of the test phase, there were no differences between participants in the MTO and OTM groups, regarding FR on the last SYM trials; while participants in the LS group had the highest FR. For the last five EQF trials, the participants in the MTO showed the highest FR of the three groups. Thereby the FR for the training phase and the first part of the test were results were also in accordance with Steingrimsdottir and Arntzen (accepted) findings, with the exception that FR was lower on SYM trials at the end of test relative to BSL trials. Concerning the different between the results from the end of the test from Steingrimsdottir and Arntzen (accepted) and the current study, could be attributed to the fact that they only used MTO training structure, while this current study compared the three training structures; and that there was possibility of establishing five 3-member classes in their study while the in the current study there was a possibility of establishing three 5-member classes. There would be five comparisons vs. three comparisons. Hence, it is a possibility that different numbers of comparisons on a display, call for different types of ocular observing response.

Transition Rate

Participants in all three groups showed an increase in TR from the last training block to the test. Especially the OTM group had the highest TR out of the three groups, on both first and last five BSL trials. TR was lowest for the participants from the OTM group; and highest for the participants from the MTO group, both in the first five and the last five SYM trials. TR for all the participants showed a reduction from the first five relative to the last five EQF trials. The LS group had the lowest TR on both first and last five EQ trials; while both MTO; and OTM had the equal TR. A possible explanation of LS training structures lack of effectiveness in establishing stimulus equivalence could be that there is a type of ocular observing response which is established, which results in shaping of maladaptive eye movements—which would be consistent with some research findings on how observing response is critical for correct visual stimulus control (e.g., Dinsmoor, 1985; Dube et al., 2010; Dube et al., 1999; Dube et al., 2003; Dube et al., 2006). Moreover, with the discrimination analysis as a premise, one would readily see that there would be a higher chance of the participants in the MTO group to "anticipate" the comparison stimuli relative to participants in the OTM group, this would explain the steep reduction in TR for the MTO group during training, as well as the OTM groups higher TR relative to that of MTO group. When visually assessing the TR of derived relations between the MTO and OTM group, the MTO group has higher TR relative to the OTM group. One interpretation of this could be again that the predictability now is in favor of the participants of the OTM group while for the participants in the MTO group it now calls for a new form of observing, hence the higher TR.

In sum the results from the current study are in accordance with the ongoing trend showing that there are minimal differences between the MTO and the OTM training structures, and that the OTM training structure is most efficient in establishing stimulus equivalence—as well as to show that there was a function of training structure on the establishment of stimulus equivalence class formation. Moreover, it also replicated earlier findings regarding RT from previous research. Although, the discrimination analysis indeed did predict several aspects of the findings, some findings were not in accordance with the discrimination analysis, which is also in line with the latest research.

Furthermore, the results did replicate some of the previous findings concerning that of ocular observing response, as well as to expand the analysis with the use of TR as a novel additional measuring unit. Future research should examine the function of training structures in relation to expansion of numbers of classes and/or class size, to assess if there are differences in the establishment of stimulus equivalence and eye movements. Likewise, in regard of Arntzen and Vaidya (2008) remarks on problemsolving behavior being different for children relative to adults, and the current study's findings it would be advantageous to see if training structures affected how stimulus equivalence was established for children relative to younger adults, and if there would be observed a difference in their ocular observing behavior. To address the question concerning maladaptive eye movements being shaped during training—considering that the LS training structures inefficiency in establishing stimulus equivalence relative to the other training structures—there would be interesting to see if an ocular observing response requirement of fixating on all comparison stimuli during the training phase could affect outcomes on test or if it would even out differences between the three training structures.

Although one possible limitation of using additional behavioral measures could be that they are treated as causal units or presuming that one can analogously consider them as how other scientific fields treat them—the plausibility of this is minimal considering the ontology and epistemology which behavior analysis is founded upon equates uncovering environment-behavior relations as its success criterion, where the cause of the behavior always is attributed to environmental variables. That is why it is critical that one is consistent, not only in utilizing eye-tracking, but also in how one conducts the research enterprise with eye-tracking—so as to get maximum benefit of the advancement of the evolving field of eye-tracking—which is in constant development of technical refinements making the research instruments even more accessible, precise, user-friendly and affordable. Moreover, at first glance one could be lead to believe that one limitation concerning this current study is the lack of exact correspondence between some of the data from the additional measures and that this would be conflicting—but the fact may as well be that by using eye-tracking and different types of additional measures, one is actually observing experimental control over behavior that before either has not been observed or has been obscured from earlier experimental analysis. This would fit perfectly in under what Sidman calls “experiments performed to try out a new method or technique,” (Sidman, 1960/1988, p. 16) this because of its success in making ocular observing responses available for analysis, to study stimulus equivalence. Indeed, the premise of the current experiment fits with the before mentioned category as well as to the category of “experiments performed to explore the conditions under which phenomenon occurs.” (Sidman, 1960/1988, p. 33) If indeed future research expand on these findings, it could have immense value as they implore further exploration, as of the nature of the different relations and the units of analysis within stimulus equivalence, which Pilgrim and Galizio (2000) generated attention towards.

The Implications regarding this research, is its advancement both to the understanding of how stimulus equivalence is established, and to the level at which the analysis is made available with the use of eye-tracking, where it gives a more detailed insight into what happens within trials. In this case is safe to presume that the utilization of eye-tracking has been fruitful and that it could have a substantial positive impact in the stimulus equivalence paradigm, which again could have an effect on applied teaching programs and interventions.

References

Arntzen, E. (2010). Stimulusekvivalens. Teoretiske betraktninger og noen praktiske implikasjoner. In S. Eikeseth & F. Svartdal (Eds.), *Anvendt atferdsanalyse 2 ed.* (pp. 100–138). Oslo: Gyldendal Akademika.

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*, 123–135. Retrieved from http://www.ejoba.org.

Arntzen, E., Braaten, L. F., Lian, T., & Eilifsen, C. (2011). Response-to-sample requirements in conditional discrimination procedures. *European Journal of Behavior Analysis, 12*, 505–522. Retrieved from http://www.ejoba.org/.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and the subsequent performance on the tests for stimulus equivalence. *The Psychological Record, 60*, 437–462. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal Of Behavior Analysis 12*, 483–503. Retrieved from http://www.ejoba.org.

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47*, 309–320. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Arntzen, E., & Holth, P. (2000). Equivalence outcome in single subjects as a function of training structure. *The Psychological Record, 50*, 603–628. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Arntzen, E., & Nikolaisen, S. L. (2011). Establishing equivalence classes in children using familiar and abstract stimuli and many-to-one and one-to-many training structures. *European Journal of Behavior Analysis, 12*, 105–120. Retrieved from http://www.ejoba.org/.

Arntzen, E., Petursson, P., Sadeghi, P., & Eilifsen, C. (2015). Equivalence class formation in accuracy or speed conditions: Immediate emergence, adduction, and retention. *The Psychological Record, 65*, 141–159. doi: 10.1007/s40732-014-0097-9

Arntzen, E., & Vaidya, M. (2008). The effect of baseline training structure on equivalence class formation in children. *Experimental Analysis of Human Behavior Bulletin, 29*, 1–8. Retrieved from http://www.eahb.org/bulletin/bulletin-archives/.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago, IL: University of Chicago Press.

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58*, 183–204. doi: 10.1901/jeab.1992.58-183

Dinsmoor, J. A. (1983). Some more information on observing and some more observations on information. *Behavioral and Brain Sciences, 6*, 718–728. doi: 10.1017/s0140525x00057976

Dinsmoor, J. A. (1985). The role of observing and attention in establishing stimulus control. *Journal of the Experimental Analysis of Behavior, 439*, 365–381. doi: 10.1901/jeab.1985.43-365

Dinsmoor, J. A., Bowe, C. A., Green, L., & Hanson, J. (1988). Information on response requirements compared with information on food density as a reinforcer of observing in pigeons. *Journal of the Experimental Analysis of Behavior, 49*, 229–237. doi: 10.1901/jeab.1988.49-229

Dube, W. V., Dickson, C. A., Balsamo, L. M., O'Donnell, K. L., Tomanari, G. Y., Farren, K. M. et al. (2010). Observing behavior and atypically restricted stimulus control. *Journal of Experimental Psychology, 94*, 297–313. doi: 10.1901/jeab.2010.94-297

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D. S., Balsamo, L. M., & Fowler, T. R. (1999). Eye tracking assessment of stimulus overselectivity in individuals with mental retardation. *Experimental Analysis of Human Behavior Bulletin, 17*, 8–14. Retrieved from http://www.eahb.org/bulletin/bulletin-archives/.

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D. S., Balsamo, L. M., Fowler, T. R. et al. (2003). Stimulus overselectivity and observing behavior in individuals with mental retardation. In S. Soraci & K. Murata-Soraci (Eds.), *Visual information processing* (pp. 109–123). Westport, CT: Praeger.

Dube, W. V., M., B. L., Fowler, T. R., Dickson, C. A., Lombard, K. M., & Tomanari, G. Y. (2006). Observing behavior topography in delayed matching to multiple samples. *The Psychological Record, 56*, 233–244. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8–12. Retrieved from http://www.eahb.org/bulletin/bulletin-archives/.

Eilifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for stimulus equivalence. *European Journal of Behavior Analysis, 10*, 187–202. Retrieved from http://www.ejoba.org.

Fields, L., Hobbie-Reeve, S. A., Adams, B. J., & Reeve, K. F. (1999). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record, 49*, 703–724. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*, 317–332. doi: 10.1901/jeab.1987.48-317.

Guinther, P. M., & Dougher, M. J. (2015). The clinical relevance of stimulus equivalence and relational frame theory in influencing the behavior of verbally competent adults. *Current Opinion in Psychology, 2*, 21–25. doi: 10.1016/j.copsyc.2015.01.015

Hayhoe, M. M. (2004). Advances in relating eye movements and cognition. *Infancy, 6*, 267–274. doi: 10.1207/s15327078in0602_7

Holland, J. G. (1957). Technique for behavioral analysis of human observing. *Science, 125*, 348–350. doi: 10.1126/science.125.3243.348

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record, 48*, 81–110. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Holth, P., & Arntzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record, 50*, 305–337. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Hove, O. (2003). Differential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. *The Psychological Record, 53*, 617–634. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*, 107–124. doi: 10.1901/jeab.2006.58-04

Johnson, C., & Sidman, M. (1993). Conditional discrimination and equivalence relations: Control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 59*, 333–347. doi: 10.1901/jeab.1993.59-333

Kirshner, H., & Sidman, M. (1972). Scanning patterns in aphasic patients during matching-to-sample tests. *Neuropsychologia, 10*, 179–184. doi: 10.1016/0028-3932(72)90057-7

Palmer, D. C. (2010). Behavior under the microscope: Increasing the resolution of our experimental procedures. *Behavior Analyst, 33*, 37–45.

Perez, W. F., Endemann, P., Pessôa, C. V. B. B., & Tomanari, G. Y. (2014). Assessing stimulus control in a discrimination task with compound stimuli: Evaluating testing procedures and tracking eye fixations. *The Psychological Record, 65*, 83–88. doi: 10.1007/s40732-014-0092-1

Pessôa, C. V. B. B., Huziwara, E. M., Perez, W. F., Endemann, P., & Tomanari, G. Y. (2009). Eye-fixation to figures in a four-choice situation with luminance balanced area. *Journal of Eye Movement Research, 2*, 1–6. Retrieved from http://www.jemr.org.

Pilgrim, C., & Galizio, M. (2000). Stimulus equivalence and unit of analysis. In J. C. Leslie & D. E. Blackman (Eds.), *Experimental and applied analysis of human behavior* (pp. 111–126). Reno, NV: Context Press.

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record, 43*, 725-744. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Saunders, R. R., Chaney, L., & Marquis, J. G. (2005). Equivalence class establishment with two, three, and four-choice matching-to-sample by senior citizens. *The Psychological Record, 55*, 539–559. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior, 71*, 195–214. doi: 10.1901/jeab.1999.71-195

Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior, 72*, 117–137. doi: 10.1901/jeab.1999.72-117

Saunders, R. R., & McEntee, J. E. (2004). Increasing the probability of stimulus equivalence with adults with mild mental retardation. *The Psychological Record, 54*, 423–435. Retrieved from http://opensiuc.lib.siu.edu/tpr/.

Schroeder, S. R. (1970). Selective eye movements to simultaneously presented stimuli during discrimination. *Perception & Psychophysics, 7*, 121–124. doi: 10.3758/BF03210147

Schroeder, S. R., & Holland, J. G. (1968a). Operant control of eye movements. *Journal of Applied Behavior Analysis, 1*, 161–166. doi: 10.1901/jaba.1968.1-161

Schroeder, S. R., & Holland, J. G. (1968b). Operant control of eye movements during human vigilance. *Science, 161*, 292–293. doi: 10.1126/science.161.3838.292

Schroeder, S. R., & Holland, J. G. (1969). Reinforcement of eye movement with concurrent schedules. *Journal of the Experimental Analysis of Behavior, 12*, 897–903. doi: 10.1901/jeab.1969.12-897

Sidman, M. (1988). *Tactics of scientific research: Evaluating experimental data in psychology.* Boston, MA: Authors Cooperative. (Original work published 1960)

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston, MA: Authors Cooperative.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5–22. doi: 10.1901/jeab.1982.37-5

Smeets, P. M., & Barnes-Holmes, D. (2005). Establishing equivalence classes in preschool children with one-to-many and many-to-one training protocols. *Behavioural Processes, 69*, 281–293. doi: 10.1016/j.beproc.2004.12.009

Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*, 643–659. doi: 10.1901/jeab.1996.65-643

Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using match-to-sample procedures: Sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities, 6*, 41–58. doi: 10.1016/0270-4684(86)90005-4

Steingrimsdottir, H. S., & Arntzen, E. (accepted). Eye-movements during conditional discrimination learning. *The Psychological Record*.

Tomanari, G. Y., Balsamo, L. M., Fowler, T. R., Lombard, K. M., Farren, K. M., & Dube, W. V. (2007). Manual and ocular observing behavior in human subjects. *European Journal of Behavior Analysis, 8*, 29–40. Retrieved from http://www.ejoba.org.

Wyckoff, L. B. (1952). The role of observing responses in discrimination learning. Part I. *Psychological Review, 59*, 431–442. doi: 10.1037/h0053932

Wyckoff, L. B. (1959). Toward a quantitative theory of secondary reinforcement. *Psychological Review, 66*, 68–78. doi: 10.1037/h0046882

Yarbus, A. L. (1967). *Eye movements and vision* (B. Haigh, Trans.). New York, N.Y.: Plenum Press.

Table 1

Results for Participants From the Linear Series Training Structure Group

Participant	Gender	Age	Training Trials	Errors	% Correct	BSL	SYM	TRA	EQ	EQ-Formation	DG 60% Block	Last Block
13003	Male	22	432	117	73	**34**	32	42	46	88	6	12
13051	Male	22	612	185	70	**34**	**33**	42	41	83	7	17
13011	Female	28	360	76	79	18	15	21	23	44	3	10
13024	Male	22	540	166	69	23	18	21	20	41	7	15
13028	Male	26	612	190	69	**33**	26	18	18	36	8	17
13043	Female	24	576	181	69	21	14	18	16	34	6	16
13033	Male	23	1116	466	58	26	12	15	17	32	14	31
13038	Male	25	360	84	77	**34**	18	17	13	30	4	10
13020	Male	24	720	202	72	24	21	13	16	29	7	20
13016	Female	20	540	155	71	**34**	28	12	15	27	3	15
Average												
		23.6	586.8	182.2	70.6	28.1	21.7	21.9	22.5	44.4	6.5	16.3

Note. The table shows participant information, number for training trails needed to complete the training phase, numbers of errors in the training phase, percent correct responding in the training phase. Numbers of correct responding in the test phase on: BSL = baseline relations; SYM = symmetry trials; TRA = transitivity trials; EQ = global equivalence; EQ-Formation = Equivalence formation, i.e. pooled transitivity and equivalence. The DG>0.60 block, shows which block was the first block were approximately 2/3 correct in one block. Last block shows number of blocks in the training phase. Numbers in bold indicate that responding was in accordance with the mastery criterion.

Table 2

Results for Participants From the Many-To-One Training Structure Group

Participant	Gender	Age	Training Trials	Errors	% Correct	BSL	SYM	EQ-Formation	DG > 60% Block	Last Block
13001	Male	25	396	124	69	**36**	**36**	**108**	4	11
13002	Male	27	396	87	78	**35**	**36**	**108**	3	11
13040	Female	25	684	236	65	**36**	**36**	**108**	10	19
13025	Male	25	540	181	66	**36**	**36**	**106**	8	15
13039	Male	28	1188	275	77	**36**	**36**	**106**	8	33
13036	Male	27	576	253	56	**36**	**36**	**104**	10	16
13017	Male	19	720	233	68	**36**	**36**	**99**	7	20
13046	Female	20	612	250	59	32	**33**	88	9	17
13019	Male	27	1080	351	68	29	29	68	10	30
13015	Female	22	324	79	76	**36**	**35**	50	3	9
Average		24.5	651.6	206.9	68.1	34.8	34.9	94.5	7.2	18.1

Note. The table shows participant information, number for training trails needed to complete the training phase, numbers of errors in the training phase, percent correct responding in the training phase. Numbers of correct responding in the test phase on: BSL = baseline relations; SYM = symmetry trials; EQ-Formation = Equivalence formation. The DG>0.60 block, shows which block was the first block were approximately 2/3 correct in one block. Last block shows number of blocks in the training phase. Numbers in bold indicate that responding was in accordance with the mastery criterion.

Table 3

Results for Participants From the One-To-Many Training Structure Group

Participant	Gender	Age	Training Trials	Errors	% Correct	BSL	SYM	EQ-Formation	DG > 60% Block	Last Block
13005	Female	24	468	179	62	**36**	**36**	**108**	7	13
13035	Male	25	396	121	69	**36**	**36**	**108**	6	11
13047	Male	21	288	69	76	**36**	**36**	**108**	3	8
13006	Male	21	324	75	77	**36**	**36**	**107**	3	9
13022	Male	29	324	78	76	**36**	**36**	**107**	3	9
13009	Male	23	396	121	69	**36**	**35**	**106**	5	11
13034	Male	24	252	60	76	**36**	**36**	**104**	3	7
13018	Female	20	1188	550	54	**34**	**35**	**103**	23	33
13053	Male	24	576	209	64	**36**	**34**	**103**	8	16
13045	Female	24	612	237	61	**36**	**35**	**101**	11	17
Average										
		23.5	482.4	169.9	68.4	35.8	35.5	105.5	7.2	13.4

Note. The table shows participant information, number for training trails needed to complete the training phase, numbers of errors in the training phase, percent correct responding in the training phase. Numbers of correct responding in the test phase on: BSL = baseline relations; SYM = symmetry trials; EQ-Formation = Equivalence formation. The DG>0.60 block, shows which block was the first block were approximately 2/3 correct in one block. Last block shows number of blocks in the training phase. Numbers in bold indicate that responding was in accordance with the mastery criterion.

Table 4

Paired Samples Statistics for Reaction Time Across Participants

		Mean	N	Standard Deviation	Standard Error Mean	*t*(29)	Sig. (2-tailed)	*d*
Pair 1	Last Training Block	1857.5833	30	473.35966	86.42325	-4.507	**.000**	-0.08
	First 5 BSL Trials	3036.1333	30	1536.48133	280.52183			
Pair 2	Last Training Block	1857.5833	30	473.35966	86.42325	-4.579	**.000**	-0.08
	First 5 SYM Trials	4337.6667	30	3089.87161	564.13079			
Pair 3	Last Training Block	1857.5833	30	473.35966	86.42325	-8.513	**.000**	-0.16
	First 5 EQF Trials	6069.0000	30	2843.54264	519.15748			
Pair 4	First 5 BSL Trials	3036.1333	30	1536.48133	280.52183	2.424	**.022**	0.44
	Last 5 BSL Trials	2309.2000	30	917.48236	167.50860			
Pair 5	First 5 SYM Trials	4337.6667	30	3089.87161	564.13079	3.456	**.002**	0.63
	Last 5 SYM Trials	2354.7000	30	800.03389	146.06554			
Pair 6	First 5 EQF Trials	6069.0000	30	2843.54264	519.15748	6.000	**.000**	1.10
	Last 5 EQF Trials	3083.0667	30	1614.16418	294.70471			

Note. The table shows a dependent-samples *t*-test across participants for reaction time data from the last block of the training phase and the different trials; and between the beginning and the end of the test phase. First 5 BSL = first five baseline trials; First 5 SYM = first five symmetry trials; First 5 EQF = first five equivalence trials; Last 5 BSL = last five baseline trials; Last 5 SYM = last five symmetry trials; Last 5 EQF = last five equivalence trials. The numbers in bold indicate significant differences. *d* = effect size.

Table 5

Paired Samples Statistics for Fixation Time Across Participants

		Mean	N	Std. Deviation	Standard Error Mean	*t*(29)	Sig. (2-tailed)	*d*
Pair 1	Last Training Block	429.4333	30	102.64995	18.74123	-2.476	**.019**	-0,45
	First 5 BSL Trials	519.0500	30	182.21569	33.26788			
Pair 2	Last Training Block	429.4333	30	102.64995	18.74123	-2.830	**.008**	-0,52
	First 5 SYM Trials	608.3000	30	353.49686	64.53940			
Pair 3	Last Training Block	429.4333	30	102.64995	18.74123	-4.639	**.000**	-0,85
	First 5 EQF Trials	624.3833	30	212.44101	38.78625			
Pair 4	First 5 BSL Trials	519.0500	30	182.21569	33.26788	1.518	.140	0,28
	Last 5 BSL Trials	456.5833	30	125.64153	22.93890			
Pair 5	First 5 SYM Trials	608.3000	30	353.49686	64.53940	1.938	.062	0,35
	Last 5 SYM Trials	475.0833	30	181.35591	33.11091			
Pair 6	First 5 EQF Trials	624.3833	30	212.44101	38.78625	2.229	**.034**	0,41
	Last 5 EQF Trials	514.2833	30	185.34807	33.83977			

Note. The table shows a dependent-samples *t*-test across participants for fixation time from the last block of the training phase and the different trials; and between the beginning and the end of the test phase. First 5 BSL = first five baseline trials; First 5 SYM = first five symmetry trials; First 5 EQF = first five equivalence trials; Last 5 BSL = last five baseline trials; Last 5 SYM = last five symmetry trials; Last 5 EQF = last five equivalence trials. The numbers in bold indicate significant differences. *d* = effect size.

Table 6

Paired Samples Statistics for Fixation Rate Across Participants

		Mean	N	Std. Deviation	Standard Error Mean	*t*(29)	Sig. (2-tailed)	*d*
Pair 1	Last Training Block	.6667	30	.40241	.07347	-3.081	**.004**	-0,5625405
	First 5 BSL Trials	1.0400	30	.72474	.13232			
Pair 2	Last Training Block	.6667	30	.40241	.07347	-3.799	**.001**	-0,6936585
	First 5 SYM Trials	1.4267	30	1.27764	.23326			
Pair 3	Last Training Block	.6667	30	.40241	.07347	-10.224	**.000**	-1,8665577
	First 5 EQF Trials	2.1400	30	1.01322	.18499			
Pair 4	First 5 BSL Trials	1.0400	30	.72474	.13232	.169	.867	0,03077875
	Last 5 BSL Trials	1.0067	30	.66950	.12223			
Pair 5	First 5 SYM Trials	1.4267	30	1.27764	.23326	2.945	**.006**	0,53765878
	Last 5 SYM Trials	.7600	30	.51835	.09464			
Pair 6	First 5 EQF Trials	2.1400	30	1.01322	.18499	5.222	**.000**	0,95334128
	Last 5 EQF Trials	1.1000	30	.81495	.14879			

Note. The table shows a dependent-samples *t*-test across participants for fixation rate data from the last block of the training phase and the different trials; and between the beginning and the end of the test phase. First 5 BSL = first five baseline trials; First 5 SYM = first five symmetry trials; First 5 EQF = first five equivalence trials; Last 5 BSL = last five baseline trials; Last 5 SYM = last five symmetry trials; Last 5 EQF = last five equivalence trials. The numbers in bold indicate significant differences. *d* = effect size.

Table 7

Paired Samples Statistics for Transition Rate Across Participants

		Mean	N	Std. Deviation	Standard Error Mean	*t*(29)	Sig. (2-tailed)	*d*
Pair 1	Last Training Block	.154	30	.1786	.0326	-2.075	**.047**	0.38
	First 5 BSL Trials	.260	30	.3114	.0569			
Pair 2	Last Training Block	.154	30	.1786	.0326	-2.388	**.024**	-0.44
	First 5 SYM Trials	.360	30	.4651	.0849			
Pair 3	Last Training Block	.154	30	.1786	.0326	-4.869	**.000**	0.89
	First 5 EQF Trials	.533	30	.4405	.0804			
Pair 4	First 5 BSL Trials	.260	30	.3114	.0569	.556	.582	0.10
	Last 5 BSL Trials	.207	30	.3769	.0688			
Pair 5	First 5 SYM Trials	.360	30	.4651	.0849	1.635	.113	0.30
	Last 5 SYM Trials	.207	30	.2852	.0521			
Pair 6	First 5 EQF Trials	.533	30	.4405	.0804	1.459	.155	0.27
	Last 5 EQF Trials	.393	30	.4017	.0733			

Note. The table shows a dependent-samples *t*-test across participants for transition rate data from the last block of the training phase and the different trials; and between the beginning and the end of the test phase. First 5 BSL = first five baseline trials; First 5 SYM = first five symmetry trials; First 5 EQF = first five equivalence trials; Last 5 BSL = last five baseline trials; Last 5 SYM = last five symmetry trials; Last 5 EQF = last five equivalence trials. The numbers in bold indicate significant differences. *d* = effect size.

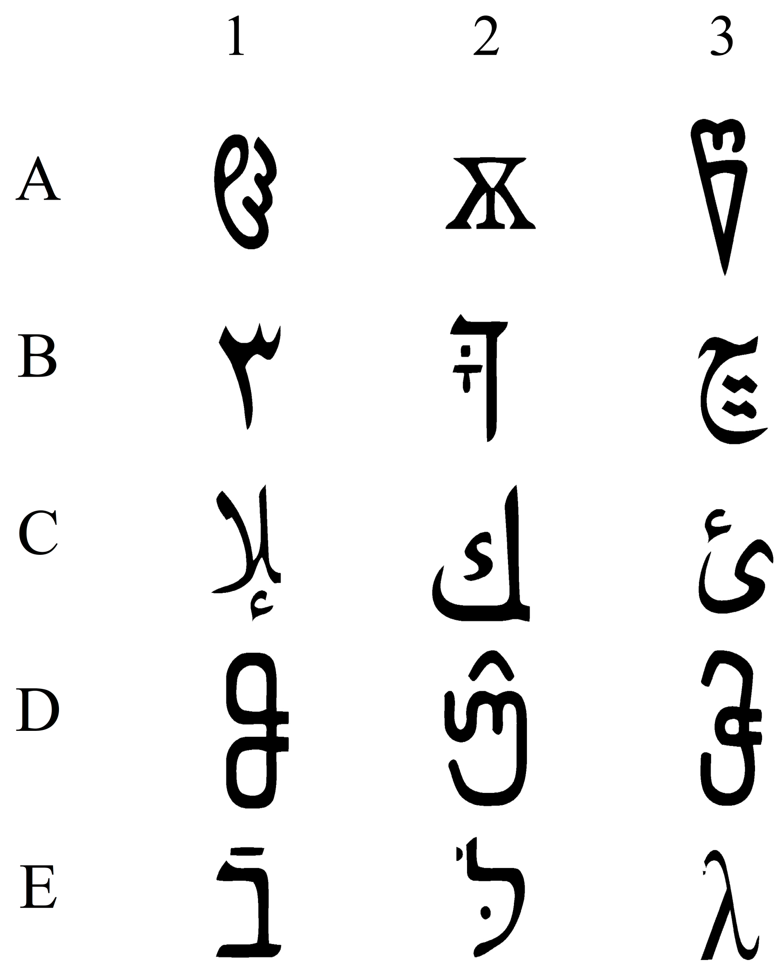

Figure 1. Stimulus set used in the experiment. The first row from the left was the Class 1; the middle row was Class 2; and the row to the right was Class 3. Letters to the left are names of the members of each respective class.

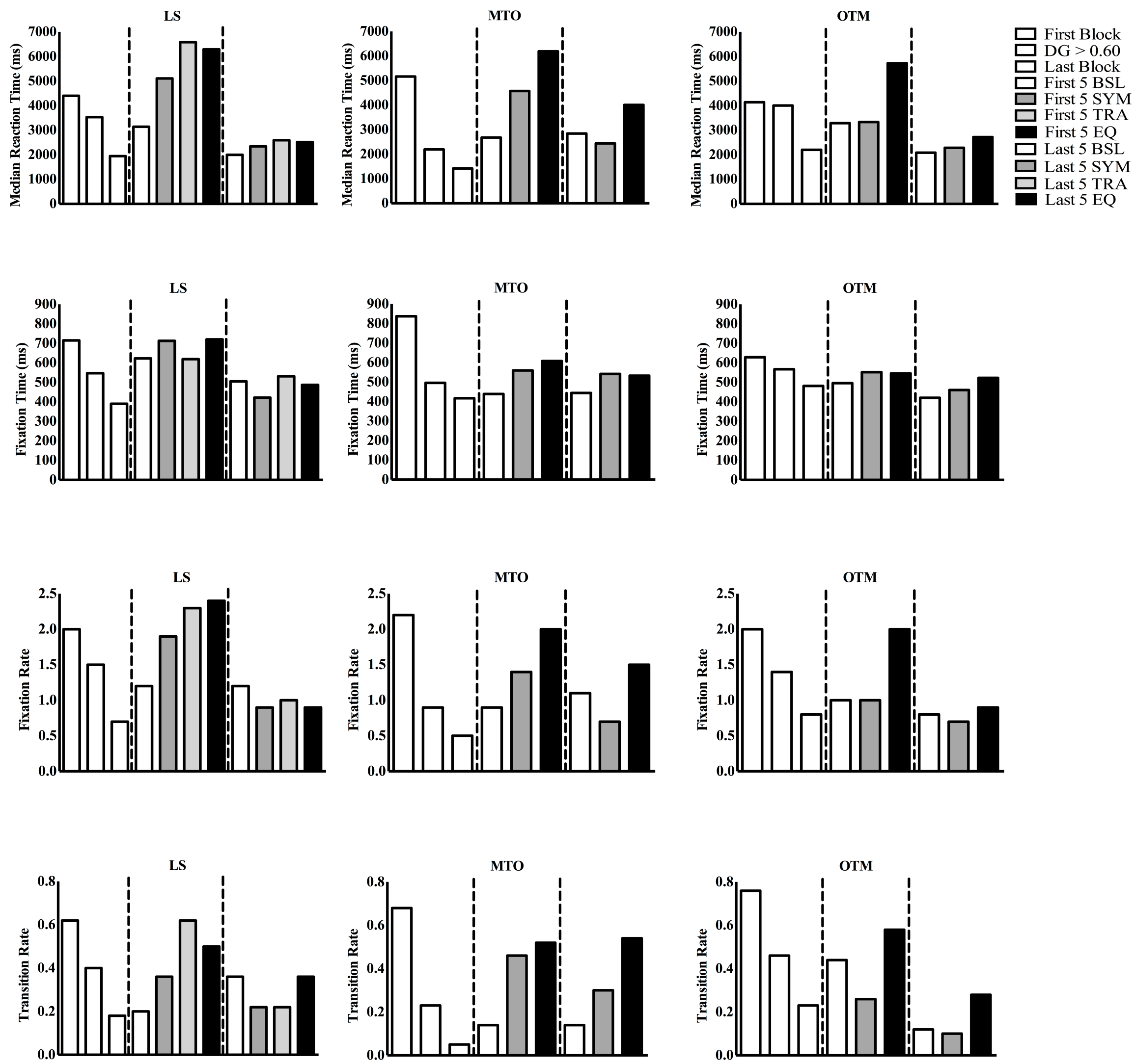

Figure 2. Average scores for each group, for each respective measure. The top panel shows median reaction for all three groups. The second top panel illustrates fixation times. The second bottom panel shows fixation rate. The bottom panel shows transition rate. First Block; DG > 0.60; Last Block; First 5 BSL = first five baseline trials; First 5 SYM = first five symmetry trials First 5 TRA = first five transitivity trials; First 5 EQ = first five equivalence trials; Last 5 DT = last five baseline trials; Last 5 SYM = last five symmetry trials; Last 5 TRA = last five transitivity trials; Last 5 EQ = last five equivalence trials.